RELATIVITY
AND
COMMON SENSE

PRINTED IN GREAT BRITAIN

PREFACE

THE scientific Theory of Relativity concerns only physical phenomena, that is to say those facts of experience which can be described completely in terms of measurements made by means of measuring rods and clocks.

The theory is no more than a way of looking at these things, but it is a very effective way. To be able to look with understanding at any complex group of things it is necessary that the things shall be arranged according to some system. In the case of very numerous things, such as the phenomena of nature, the arrangement has to be in the form of a written record, preferably in the language of mathematics—because this is the most concise and the least ambiguous of languages and is better suited than any other to logical reasoning. The system of the record is determined by the factors chosen for the description of the phenomena.

The importance of the choice of factors may be illustrated in reference to two possible ways of looking at the view which a man sees through a glass window. If he considers that what he "really" sees is the glass of his window and the peculiar changes its surface undergoes (changes which, from the ordinary point of view are seen to belong "really" to the landscape beyond the window) he will naturally make his description of the picture in terms of changes occurring on the surface of the glass—a 2-dimensional space continuum. In fact he will do just what the landscape painter tries to do (or, if he be ultra-modern, refuses to do), namely, describe the landscape in terms of the exactly measured position of patches of light and shade upon the surface of the glass.

He will find that the description required is a changing one demanding a kinematographic series of 2-dimensional pictures arranged in uniform time-succession.

If now he has to describe the picture in words or other conventional symbols he will—if he is worldly wise—give up art and become a physicist. He will then have a reasonable chance of making an unambiguous description of the landscape, because he will use the physicist's arbitrarily chosen and eminently worldly definition of "reality."

He will find factors of description ready to hand, namely particles and their space and time relations. These factors will be indicated by brief symbols and their relations expressed by means of algebraic equations. Thus, for instance, the portrait of a primitive man given in Fig. A now becomes the simple equation

$$y = 4 \sin x + \sin 5x.$$

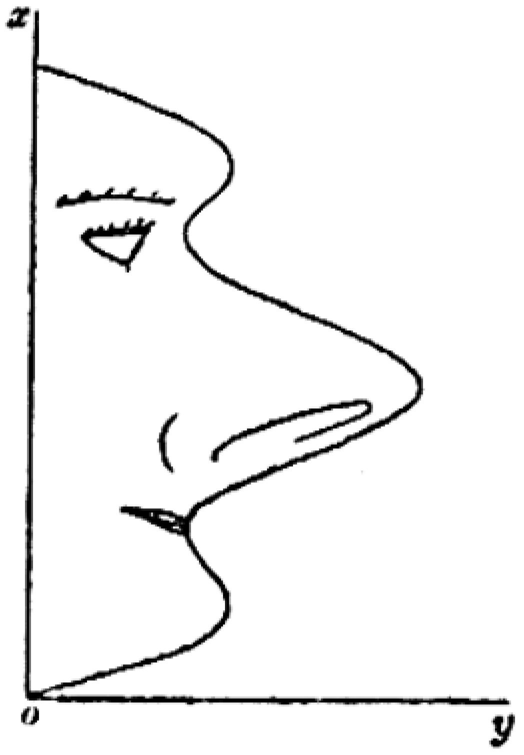

FIG. A

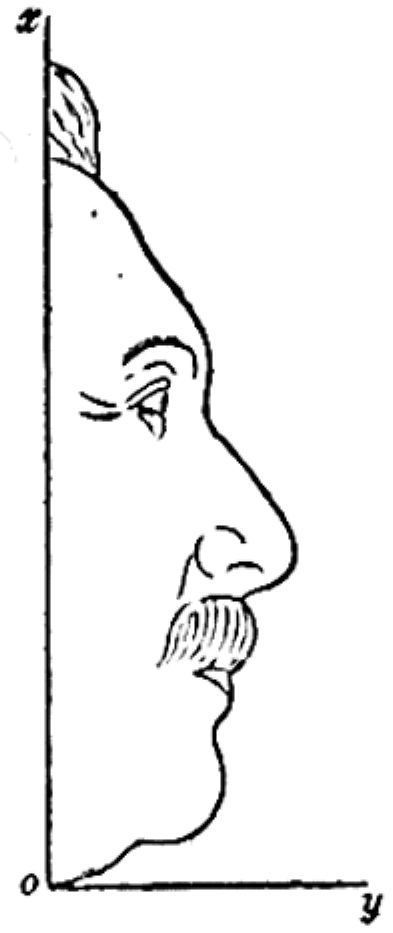

FIG. B

This equation describes the positions on the paper of all the little black particles which compose the profile.

The thoughtful intellectual profile of Fig. B (suggested by a recent picture of Dr Einstein*) demands an equation having over 100 terms, namely:

$$\begin{aligned} y = A_1 \sin x + A_3 \sin 3x + A_5 \sin 5x + \ldots \\ + A_{(2n-1)} \sin (2n-1)\, x + B_1 \cos x + B_3 \cos 3x \\ + B_5 \cos 5x + \ldots + B_{(2n-1)} \cos (2n-1)\, x, \end{aligned}$$

* The profile is derived from the portrait (permission to use which was very kindly given both by Dr Einstein and by the publishers of the journal) published in the *Illustrirte Zeitung* (Berlin) of 23rd April, 1922.

where the A's and the B's are numerical quantities and n should be greater than 50.

Other additional equations could be given to express the eye and lips and wrinkles.

The group of equations thus obtained expresses all that there is of scientific "reality" in the portrait—for it expresses that and only that in the portrait upon which all observers must agree. "Beauty" is not a scientific reality because experience shows that beauty is something imputed to an object by an observer: it may be seen by one observer and not by another, although there is no generally acceptable criterion as to which of the two observers' visions is the more "true."

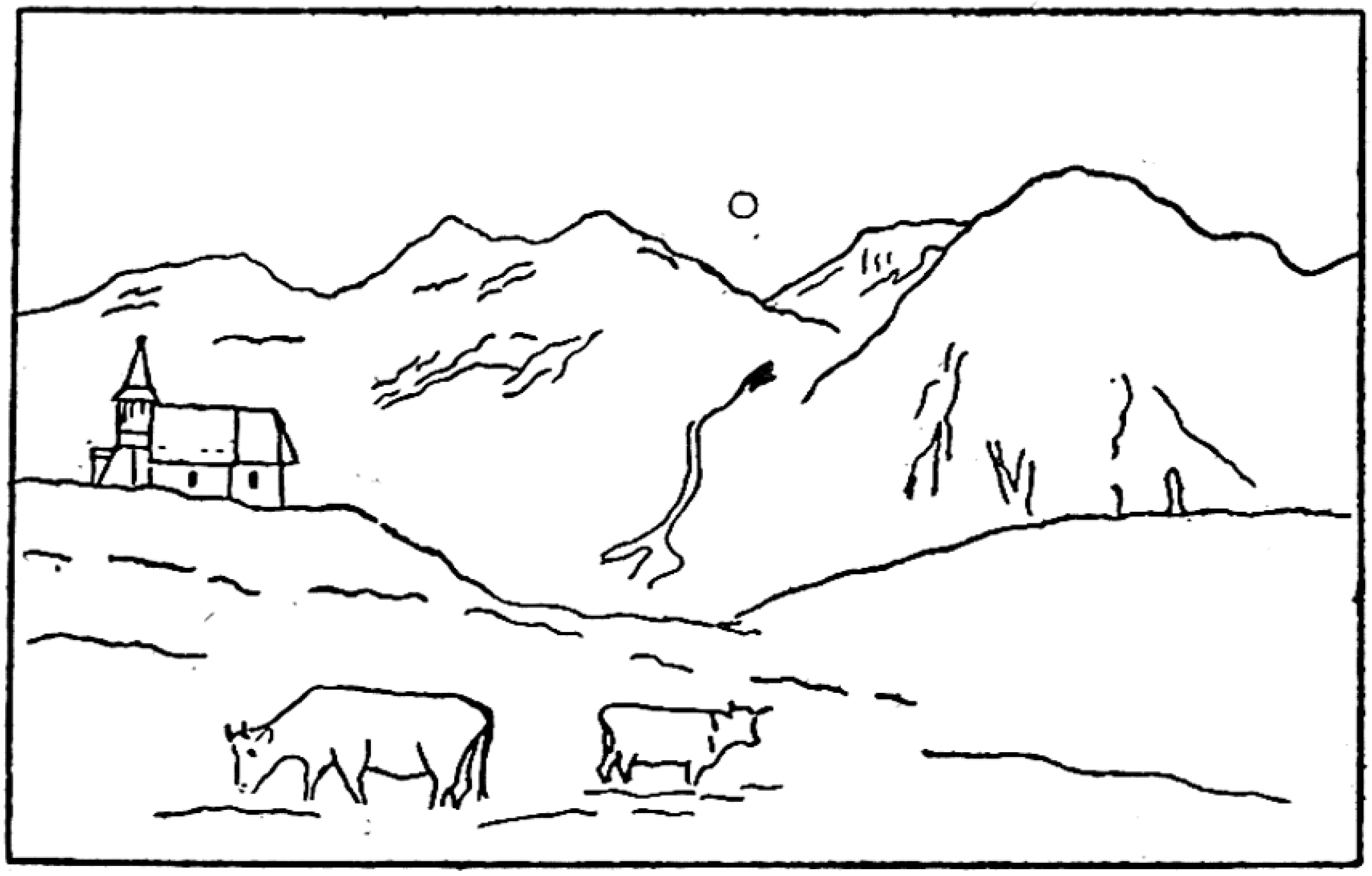

FIG. C

In this man's description of the landscape the mountain-range forming the horizon (Fig. C) will be represented by some such equation as that of the portrait, while the rising moon, being "full," will be described by the familiar equation to a circle.

The cows in the foreground will demand equations so complex as to make the physicist long to be a mere artist again, with the artist's freedom to cover by a "careless rapture" anything which he cannot easily describe; nevertheless, the equations to these cows are possible and unambiguous. The chief difficulty with the cows is to represent their motion, for not only does their position on the glass (or canvas) surface change, but the size of a cow's outline is a function of the cow's position, while its shape varies without apparent cause. Sometimes on reaching the neighbourhood of another cow or of the church, one cow partially or totally disappears—sometimes, indeed, they all suffer annihilation by the church.

Thus at every instant a new set of bovine equations is required and the array of equations necessary for the continuous description of the landscape, including the cows, is unlimited.

Our physicist must search for some improved method of description if the mathematical system of recording is to be practicable—and no better system is known.

This improved method must eliminate the faults just noticed, namely,

(1) That passage of time is indicated by a halting transition from one picture (or its set of equations) to the next, and only by means of an infinitely large number of pictures can the continuous nature of time be described.

(2) The use of space in its capacity only as a 2-dimensional continuum (the surface of the glass or canvas) as a factor in the description introduces the difficulty of the apparently uncaused annihilation and re-creation of cows.

Careful consideration shows the observer that he must look at the landscape in a different way, namely, in such a way that the two factors time and the third dimension of space make their appearance, with their proper character of continuity, in the description. Thus, instead of a disconnected set of equations in x and y, one set belonging to each instant of time, the description must consist of one set of equations in x, y, z and t, where z represents the picture-depth dimension of space, and t

represents time. The way of looking at things must not be that of the stationary, one-eyed man, but rather that of the man who surveys the landscape from many different positions. Without such a survey much of the meaning of the landscape is lost, and unnecessary, mysterious factors are introduced.

The illustration which has been given is intended to bring out the importance of one's way of looking at things. To take a restricted or biassed view is to introduce unnecessary, mysterious factors into the description and so conceal much of its meaning.

The modern Theory of Relativity—as developed by Einstein—takes a point of view which is broader and less biassed than any before taken and the striking thing about this view-point is that, in addition to removing the mystery from phenomena which before had seemed mysteriously inconsistent with all other phenomena, it actually reduces the number and complexity of the factors in terms of which scientific descriptions are made. Thus mass, space, time, force, gravitation and the old laws of mechanics, have been reduced to a simple system of space-time intervals between "point-events," and their 4-dimensional geometry.

Among the hitherto inexplicable phenomena for which Relativity has found a logical place in physical science are the following:

(1) The "absurd" fact that the velocity with which a ray of light overtakes a man is always 186,000 miles per second, no matter how fast the man himself may be retreating from the source of light.

(2) The fact that the path of a light ray is affected by gravitation in exactly the same way as the path of a bullet possessing light's velocity.

(3) The fact that the elliptical orbits of the sun's planets rotate relatively to one another.

(4) The fact that the mass of a particle varies with the particle's velocity.

(5) Gravitation.

In addition to finding a logical place in science for such facts as these, Relativity has removed the anomalies which—though convention usually camouflaged them—always had existed even in the orthodox explanations of simple mechanical phenomena. In particular, by depriving the quantity called "force" of its supposed "reality," the most outstanding anomaly of mechanics has been removed, namely, the impossibility of defining force except in terms of mass, while, at the same time, mass itself had to be defined in terms of force.

The complete Theory of Relativity is called the General Theory; it includes the earlier theory—called Restricted or Special—as a special case.

The General Theory has proved already to be a remarkably effective way of looking at all the so-called mechanical phenomena as well as many electrical ones—of which optical phenomena are a special case—but the great task now before Einstein and Weyl and other Relativists is that of bringing the remaining electrical phenomena into the theory; a task which, from the well warranted point of view that all physical phenomena are essentially electrical, is of crucial importance.

Should this task prove possible—as it surely promises to do—then we shall be able to say that the General Theory of Relativity has made of Physical Science a system possessing all the logic and consistency of pure geometry. Physical Science will remain, of course, an experimental science; indeed, though the fact often is obscured, geometry itself is based upon experiment. Failure, on the other hand, to bring electrical phenomena within the scope of the General Theory in its present form would not mean the rejection of the theory, but merely would necessitate its further development, for there is, at present, no reason to think that any of the known phenomena of nature are necessarily inconsistent with the theory. Nothing would have to be changed in Einstein's General Theory except its name—which would have to be transferred to a still more general developed theory.

* * * *

Einstein's Theory of Relativity is presented in this book from the point of view taken in *Space, Time and Gravitation* (Camb. Univ. Press, 1920) by Prof. Eddington, to whom thanks are tendered for much education and much pleasure. The author wishes to thank also, Einstein, Silberstein, Lorentz, Cunningham, Schlick, Freundlich, Whitehead, Weyl and Karl Bollert.

F.M.D.

September 1924

CONTENTS

PREFACE v

INTRODUCTION 1

Popular interest in the Theory of Relativity 1
The disappointment provided 2
The object of this book 5
The "absurd" phenomena on which the theory is based . 6
Mental repugnance 9
The "true" expression for relative velocity 10
The standpoint of Relativity 11
Can the mind do without an ether? 12
Why not retain the Newtonian ether? 13
We must take things as we find them 15
Is there an "ether" of any sort? 16
Fundamentals never can be defined 17
Is there absolute motion in the "ether of events"? . . 18

CHAPTER ONE

The search after truth 20
Time intervals are unknowable 22
What is reality? 24
Definition of reality 24
The space-time continuum 25
Einstein's guess 26
Ghosts 27
The universal observer 28
What is truth? 29
Solipsism 30

CHAPTER TWO

The problems of life 31
Matter and electrons 31
Indefinable fundamentals 31
The meaning of "consistent" 32
Truth which is deeper than that of science 33
Scientific knowledge is of relations 34
The nature of explanations 34
Brevity 36
The objects of explanations 38
Familiar conceptions 39
Action at a distance 40
Possibility of mental conception no criterion of the value of an explanation 40

CHAPTER TWO (*cont.*)

What is "contact"? 42
Atomicity or continuity 44
What is a conception? 45
The usefulness of "impossible conceptions" 46
Minkowski's "world" 47
The General Theory is an "impossible conception" . . 47
The mixing of space and time 48
The difficulty of the idea of a 4-dimensional continuum . . 48
A brick needs time in which to be a brick 51
The gradient of events 52
What then is the meaning of Minkowski's 4-dimensional continuum? 52
The space-time "interval" 53
"Imaginary" time 53
Homogeneity of continua 54
Four-dimensional space 57
Examples of mixed continua 57
Choppy space and smooth space 59
The language of intervals 59
Prediction of the future 60

CHAPTER THREE

The demand for Relativity 61
The laws of Newtonian mechanics 62
Centrifugal force 63
Newton's rotating bucket 63
Absolute rotation 65
Static forces 67
Relativity discards force 69
The problem of the two apples 72
Can a body move two ways at once? 73
Each man has a right to his own special ether . . . 77

CHAPTER FOUR

The General Theory of Relativity; its point of view . . 78
The falling lift 78
Shape and size 79
Does time flow uniformly? 81
The mathematics of molluscs 83
Symbolical records 87
World-lines 89
The mollusc 89

Can a particle be observed? 90
Man as human being and man as observer 91
Space-time coincidences 91
Intersections of world-lines 94
The problem before the Relativist 96
The mathematical method 97
Difficulty of discussing Relativity 100
"Forces" are the Relativist's clue 101
Circular arguments 101

CHAPTER FIVE

Gravitation cannot be explained away 103
Centrifugal force can be explained away 104
The meaning of "explained away" 106
The problem of the reality underlying gravitation 107
The "straightening" of curves by choice of frame of reference 112
Straightening the paths of falling bodies 115
The Relativist's dilemma 117

CHAPTER SIX

The beetle 119
Hummocks in space 123
Non-Euclidean space 123
The Relativist's problem summarized 127
Mesh-systems 130
Introduction to the "g's" 134
Gradient of the "g's" 139

CHAPTER SEVEN

Gradient of the gradient of the "g's" 143
Meaning and need of the ten "g's" 150
Importance of the "g's" to Relativity 151
The ten "g's" of space-time 153
Newton's law of gravitation a special case of Einstein's . 155

CHAPTER EIGHT

The laws of space-time hummocks 157
Relativity and matter 157
Coincidences in space-time essential factors in reality . . 158
Einstein's law of gravitation 161
How had Newton erred? 162
Equation of Einstein's law of gravitation 168

CHAPTER NINE

Maps and diagrams 175
Lengths and intervals 177
Interval maps 179
The root of minus one 180
The invariant property of interval 181
"Imaginary" intervals 181
The velocity of light called unity for convenience . . . 183
Minkowski's "imaginary time" 185

CHAPTER TEN

The action of fields of force on light 186
Ambiguity of the word "straight" 190
Contraction of bodies in a gravitational field . . . 192
The Principle of Equivalence 193
Is the contraction "real"? 196
The inadequacy of explanations of Relativity in ordinary language 200

CHAPTER ELEVEN

The Special or Restricted Theory 201
Assumption of constancy of the velocity of light . . . 202
Measuring rods and clocks 204
The rod-clock 204
Reasons for expecting the constancy of the velocity of light . 208
The TOX diagram 209
Different observers use different frames 211
Change of length of a moving rod 212
Changes in units due to motion 212
The FitzGerald contraction and the TOX diagram . . . 217
The relative setting of clocks 218
Time order may be different to different observers . . . 223
The FitzGerald contraction in the light of common-sense . 224
Transverse changes of length cannot occur 224
Ether methods of explanation 232
Intervals and distances compared 234
Hyperbolic geometry 236
The Lorentz Transformation 237
Correct formula for relative velocity 238
Deduction from the Lorentz Transformation of the effect of gravitational fields on rods and on clocks 244
Gravitational potential 246
Behaviour of measuring rods in a gravitational field . . 247

CHAPTER TWELVE

Curvilinear motion 249
Path of a particle in space-time 250
The path of a particle must be the longest possible . . . 251
Semi-Euclidean geometry 253
Return to the General Theory 253
The "first law of motion" of Relativity 255
The spiral paths of planets 255
The geometry of the practical world is non-Euclidean . . 255
The geometry of any "given region" is ever changing . . 255
Behaviour of light in a permanent gravitational field . . 259
Refraction of light in a gravitational field 260

CHAPTER THIRTEEN

Matter 261
Mass of the earth in kilometre measure 264
Impenetrability of matter 264
Conservation of mass and of energy 264
The new definition of momentum 268
The new meaning of momentum 269
Energy and mass 269
Variation of mass with velocity 270
No velocity can exceed 3×10^{10} *cms. per second* . . . 270
The motion of the perihelion of Mercury 271
Non-Euclidean space 273
Outline of a method of deducing Einstein's law of gravitation from the Lorentz Transformation 274
Is space infinite? 276
The future development of Relativity 278

RELATIVITY & COMMON SENSE

INTRODUCTION

Popular interest taken in the Theory of Relativity

As a rule scientific theories are left to the experts, the ordinary man being content to steal what profit he can from their application but possessing seldom enough interest in the theories, or in their inventors, to wish either to understand the former or to reward the latter.

How comes it, then, that the Theory of Relativity has aroused such universal interest? There never was a theory which gave so little promise of commercial application, nor was there ever one so difficult to understand.

The ground of this popular interest may perhaps be disclosed by a consideration of relevant newspaper headlines:—"Light caught bending," "Newton superseded," "The 4-dimensional world," "Euclid wrong," "Is space crumpled?"

Ordinary people are not scientists but they hold firmly to a few scientific notions which early teaching as well as practical experience tell them are true. Theories such as "The Atomic Theory," "The Kinetic Theory of Gases," "The Conservation of Energy," "The Electron Theory," not one of these has threatened to interfere with the foundations of practical belief in the way that the "Theory of Relativity" presumes to do, nor has any of them provided such novel paradoxes for discussion.

A similar sort of interest is felt in any new religion whose doctrines appear unorthodox, but a new scientific theory has the great advantage over a new religion that there is no danger of losing caste (to say nothing of being damned) in venturing to discuss it with an open mind; moreover as "common-sense" notions are held with greater confidence than orthodox religious notions—the former seeming to be forced upon us willy-nilly by experience, whilst the retention of the latter seems possible only by aid of a highly cultivated faculty of "belief"—one can carry on a wordy warfare with more unfeigned vigour and earnestness in defence of common-sense than in defence of common dogma.

Thus the ground of the popularity of the Theory of Relativity seems to lie in the novelty and in the unorthodox nature—amounting apparently to paradox—of its propositions: propositions which concern not mysterious scientific things like atomic weight, or valency or entropy or potential, but the ordinary familiar things of life such as space, time, force, gravitation, light and the axioms of geometry; things about which every man feels that he is entitled to speak as confidently as any man.

The disappointment provided

However, the popular interest shown in the theory has led chiefly to confusion and disappointment. People have asked for an "explanation" and many have been offered, but not one has given satisfaction. A popular explanation means one that is expressed in ordinary common-sense terms, whereas the very basis of Einstein's theory lies in phenomena which are outside the range of common experience and therefore of common-sense. Our common-sense notions of space and time and things are simply wrong, for it is now known that the world of experience—the real world observed by all observers—contains no factors which correspond strictly to "common-sense" notions. Modern scientific experience has gone far beyond the range of common experience and has shown common-sense to be at fault.

Past experience has led to the formulation of rules, and men have made the convenient though unwarranted assumption that the proven applicability of these rules to the events of past experience guarantees their applicability to the events of new experience—as though it were the rules which determined the events. There is, for instance, the rule for finding relative velocity: if a man runs away at 8 miles per hour and his pursuer follows at 6 m.p.h. the relative velocity of the two men is 2 m.p.h. The simple arithmetic rule used in such calculations gives results so precisely in accord with all the events of *common* experience that men have come to regard the rule not as an empirical rule but as a necessary truth. A conception of space and of time which would make such a rule necessarily true is the common Newtonian conception, namely that of 3-dimensional space in which distances are absolute—so that, for instance, the length of a rigid rod is truly the same no matter in what direction or in what locality in space the rod be held or in what manner the rod may be moving—and of time, a thing independent of space and "flowing evenly," so that, for instance, a definite moment in time may be used simultaneously and is understood similarly, by every observer, no matter what may be his location and his motion in space. Common-sense has adopted such a space and such a time as of necessity "true," and with full confidence in the legitimacy of the assumption involved, it rejects as "absurd" or at least mistaken, evidence of phenomena which fail to adapt themselves to these peculiar sorts of space and time. The result of the fight between common-sense and facts is that common-sense has had to be discredited.

Experimental evidence which it has been found impossible to discredit shows, for instance, that a ray of light goes out from its source with a velocity of 186,000 miles per second ($= 3 \times 10^{10}$ cms. per sec.) relatively to the source: it shows also that, under proper conditions, an electron may be shot out at the same instant from the same light source with a velocity of, say, 80 % of light's velocity relative to the source. Common-sense demands that the velocity of that electron relative to the wave front of the ray of light shall be ($100 - 80 =$) 20 % of light's velocity or 37,200 m.p.s. Actual measurement proves that the relative velocity is not 37,200 m.p.s. but is 186,000 m.p.s.—an "obviously absurd" result.

Or again, to take an example in which no such intangible thing as light is involved, a particle *P* (Fig. 1) of radioactive substance is known to shoot off electrons—that is to say massive material particles—in all directions with very great velocities. At a given instant in time it is possible for two diametrically opposite electrons *A* and *B* to be shot out in opposite directions each with 0·8 of the velocity of light.

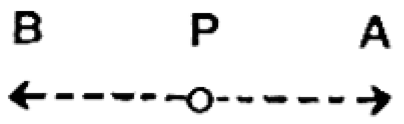

FIG. 1

Common-sense says the relative velocity of *A* and *B* is ($2 \times 0{\cdot}8 =$) 1·6 times the velocity of light; but common-sense is wrong, for actual measurement shows that the relative velocity of *A* and *B* is less than the velocity of light.

Some people prefer to say that "in reality" the relative velocity *must be* 1·6 times light's velocity although measurement shows that it *apparently* is less. The proverb is a useful one which says things are not always what they seem, but it is an unscientific one. Physical science observes and records evidence and it knows only one sort of evidence, namely data of measurement. To science that which can be measured is real and the results of measurement are facts; if there be some reality deeper than that which measurement discloses it is a sort of reality with which physical science does not concern itself.

We may, if we like, blame the instruments with which the "absurd" measure results were obtained, and may say that only such instruments shall be used as will give the results expected by "common-sense." But to do this would be to make experimental science absurdly futile and obviously unnecessary. The scientist's object is to find out facts, not to force facts into agreement with preconceived notions.

Scientists regard it as having been conclusively established that the ordinary rule for finding relative velocities is wrong. When the velocities concerned are small compared with 186,000 m.p.s. the error introduced by using the ordinary rule is negligible—and, usually, undetectable—but for high velocities the rule gives results which are inconsistent with experiment.

The acceptance of this position involves the realization that space and time are not "Newtonian." The German mathematician Minkowski was the first to realize in what manner the Newtonian ideas of space and time would have to be modified. He found that space and time could be allowed no longer to hold mutual independence; to neither space nor time could "reality" be assigned, the reality underlying our world being a 4-dimensional continuum of space and time.

Minkowski's treatment dealt, however, only with special cases; it remained for Einstein to develop and perfect what Minkowski and others had begun.

Minkowski's conception of space-time removed the apparent paradox of relative velocity but it threw no new light upon "force," or upon "gravitation." Einstein's conception of space-time, expressed in his *General Theory of Relativity*, takes force into account and leads to a new law of gravitation which promises to embody the laws of all physical phenomena.

The object of this book

The object of this book is to explain the meaning of the theory of relativity—not in "common-sense" terms, since this is impossible, but in reasonable terms such as may be followed by anyone who knows a little science and a little mathematics.

The theory can hardly have lasting interest for anyone who is entirely without scientific knowledge, and there are many possessing such knowledge whose first efforts to tackle the subject have brought discouragement—natural common-sense having been too severely shocked by the theory's "absurd" demands.

It is hoped that the method of attack here followed may succeed in disclosing only new interest in those difficulties which usually have provided discouragement and exasperation.

One word may be said in anticipation: The theory takes perturbing liberties with space and time and mass, and one begins to fear lest all the realities of practical life may be explained away. It will be found, however, that one very practical common-sense sort of thing is left untouched, namely, a coincidence in space and time.

For example, to be hit on the head by a brick is a fact—an unambiguous fact. Various observers, according to their states of motion relative to me, may have, perhaps legitimately, various notions as to my height or even the amount of matter in me, but as to my being hit on the head, not even a relativist will admit of any ambiguity about that,—by no conceivable observer can the event be interpreted as other than a coincidence in space and time: I am not asked to have patience with any observer who tries to persuade me that "in truth" the brick just missed my head, whether by an inch or by a second; any observer who says this is simply wrong.

The "absurd" phenomena on which the theory is based

The famous Michelson-Morley experiment is often named as the experiment whose explanation demanded the Theory of Relativity. Actually, however, the case is rather that the M.-M. experiment happens to be one of the more striking and insistent of a large number of experiments both in optics and in electrical science, all of which called for some new theory and all of which have proved consistent with Relativity.

In principle the M.-M. experiment consisted in sending, from a light source fixed relatively to the earth, two mutually perpendicular horizontal beams of light and comparing the times taken for them to travel each a distance of one yard. The experiment was repeated many times and at various seasons of the year but always the result was the same, namely, that the times taken by the two beams of light in travelling equal measured distances were the same.

At first sight there seems to be nothing striking—certainly nothing *absurd*—about such a result, but consideration discloses difficulties. If the experiments had concerned sound instead of light, they would have been taken to indicate that the atmosphere was permanently windless, that is to say that the air is carried round with the earth in its rotation and that local disturbances (winds) never occur.

By analogy, then, the M.-M. experiment seems to show that the light-conveying medium (the ether) is carried round with the earth and suffers no local disturbances.

Accepting such a conclusion, tentatively, how could we answer the question what is the radial thickness of the earth's stagnant blanket of ether?

Since light reaches us from stars many thousands of millions of miles distant we might be inclined to say the earth's ether blanket must be at least many millions of miles thick. In this case, however, we should have to suppose that the earth was unique among heavenly bodies in the possession of a stagnant ether. We could then speak appropriately of the ether as belonging to the earth and could say that all other heavenly bodies floated and rotated in the earth's infinite coat of ether. Such special treatment by the creator of the universe, of this one little planet seems, however, highly improbable, and it is not surprising to learn that the earth's possession of such an ether blanket is inconsistent with the observed phenomena of light reaching us from the stars. For instance, the various observed effects due to the phenomenon of astronomical aberration, on account of which in order to look at a star which is vertically overhead one must tilt the telescope a little in the direction of the earth's motion—just as, to meet a vertical shower of rain a running man must tilt his umbrella forward—would be most difficult to reconcile, quantitatively, with the existence of such an ether blanket.

If the alternative suggestion were made that the earth and every other heavenly body has its own fixed blanket of perhaps only a few miles thickness, it would then be necessary to suppose that the whole of the space in which these blanketed bodies moved was full of ether which must either be at rest relative to the blankets or moving in some way independently of the motion of the blankets. Such a notion would involve an interference with the transmission of rays of light when these passed from the interstellar ether into one of the moving blankets, such as appears contrary to all experimental evidence. And if we imagine an observer standing on the outer surface of an ether blanket we have to suppose that he would experience an ether wind, due to the relative motion of his own ether and that of interstellar space. The possibility of such an ether wind would be hard to reconcile with the fact that all experimental evidence is against the occurrence of relative motion of adjacent parts of the ether, there being indeed no evidence to prove that the ether possesses parts.

It has proved in fact impossible—if the description of natural phenomena is to be saved from weird complexity—to retain the notion of an ether fixed to the earth.

The least unsuccessful attempt to explain phenomena in terms of an all-pervading ether, is that of the Dutch physicist Lorentz, who has supposed that the universe is immersed—and soaked—in a uniform stagnant ether which is fixed relatively to the so-called "fixed stars." Through such an ether matter then moves with perfect freedom and the practically inconceivable property has to be assigned to this ether that it exerts force upon matter while matter exerts no force upon it—there is action without reaction—and the saving of Newtonian space and time by retention of the ether is seen to involve the sacrifice of one of the Newtonian laws.

Overlooking these difficulties however, we find in the ether of Lorentz a very useful thing, enabling us, if we are careful and remember always that the conception is fundamentally at fault and if carried too far will lead to wrong conclusions, to picture the phenomena of what is "really" a 4-dimensional world, in terms of 3-dimensional space and 1-dimensional time.

In regard, then, to the M.-M. experiment the position was one which appeared to involve contradiction. The experiment said there was no ether wind—since such a wind would have caused light to travel with greater velocity in the wind's direction than in other directions—while the only ether theory which could be made consistent with other phenomena (namely Lorentz's theory) said the earth did indeed move through a stagnant ether and such motion must involve the existence of an ether wind.

FitzGerald and Lorentz suggested what appeared to be the only reasonable way out of the dilemma. Experiment disclosed no wind,—ether theory said there must be a wind. The mind demanded an ether, therefore there must be an ether and an ether wind. As experiment said there was no such wind experiment must be at fault. It must be that the ether wind affected the measuring rod used in the experiments. The M.-M. experiment could then be reconciled with the idea of the ether if it were assumed that all bodies, including, of course, measuring-rods, suffered contraction in the direction of their motion whenever they moved relatively to the ether.

This suggestion of FitzGerald and Lorentz was startling but could not be proved absurd, and when, on calculation, it was shown that the change in the *relative* lengths of two bodies, one of which was moving relatively to the other (and therefore faster than the other relatively to the ether) with a velocity far exceeding velocities ordinarily dealt with in practice, was so minute as to be negligible, the suggestion became less repugnant to the mind.

Mental repugnance

Mental repugnance—in spite of its value in the affairs of ordinary life, in which, acting as an acquired intuition, it warns us against foolish actions—is apt to retard scientific progress. It was only because of the mind's repugnance against having no medium for the transmission of light, that the ether was invented, and now that the mind has been given an ether, still greater ingenuity is required in presenting to the mind the consequences of this ether in order that these consequences shall not appear repugnant.

The M.-M. experiment showed that the measured velocity of light was the same no matter in what direction, relative to the ether wind, the light might be travelling. This means, for instance, that if there were a source of light at rest in the ether, and a number of observers, some at rest relatively to the source (and therefore to the ether), others moving towards the source, and still others moving away from it, all these observers would obtain the same value for the measured velocity of light relatively to themselves. In order to measure the velocity it is assumed that all the observers use rods and clocks which, when brought together and examined while at rest in one spot, prove to be in exact agreement with standards. In principle what each observer does is to measure by his own clock the time taken for the light to travel from end to end of his own rod.

It is this conclusion, that light travels with the same velocity relative to a system retreating from the source as relative to one approaching the source, which makes the results of the

M.-M. experiment seem "absurd." The velocity with which two trains pass each other is obviously greater when they are moving in opposite directions than when one is overtaking the other, for if the one were making 50 m.p.h. and the other 40 the relative velocity in the former case would be 90 and in the latter only 10 m.p.h. To say that the relative velocity would be in each case the same is simply absurd.

The only apparent difference between this case of one train passing another and that of light passing a moving observer is in the "absolute" magnitudes of the velocities concerned, and the mind feels that the calculation of the relative velocity of the trains must be unaffected by the absolute magnitudes of the two component velocities. If instead of 50 m.p.h. and 40 m.p.h. the actual velocities had been 50 million m.p.h. and 40 million m.p.h. the answers should still have been 90 and 10, but this time in millions of miles instead of miles, per hour.

This is what the mind feels; this indeed is plain common-sense; it might be true but it is not, and common-sense is wrong.

The simple arithmetic rule commonly used for finding the relative velocity of two particles moving uniformly in the same straight line is the special form, approximately correct when the component velocities are small, of the general rule applicable to velocities of whatever magnitude, namely:

The true expression for relative velocity

If P has velocity u and P' has velocity w in the same direction, the relative velocity of P and P' is $(w-u)/(1-wu/c^2)$ where c is the velocity of light, instead of $(w-u)$ as in Newtonian mechanics. In the system of units in which the velocity of light is unity the equation for relative velocity reads:

$$\text{Relative velocity} = (w-u)/(1-wu).$$

The rule is not applicable to relative velocities greater than the velocity of light because there are no such velocities. According to the Theory of Relativity no particle of matter can travel relatively to another particle with a velocity greater than

186,000 m.p.s. If the human mind finds such an apparently arbitrary limitation of nature's possibilities absurd, it can only be because when the mind thinks of a "particle in motion," it is refusing to be limited to the real measurable particle in real measurable motion, of which alone we have experience, and is enthroning as real or absolute, mental abstractions which, though based on the realities of nature, have no exact counterpart in nature.

The standpoint of Relativity

Consideration of the experiments of which that of Michelson and Morley is typical, suggests that the phenomena of nature have been laid out with the most careful intention to prevent anyone from measuring or even detecting motion relative to the ether. Man has been allowed to invent the ether but that is all he can do in regard to it!

The ether is rigid yet allows all matter to move about in it without friction or resistance: it is elastic but cannot be distorted: it moves but its motion cannot be detected; it exerts force on matter but matter exerts no force on it; it has no mass nor has it any parts which can be identified; it is said to be at rest relatively to the "fixed stars," yet these stars are known to be in motion relatively to one another.

The standpoint of Relativity is that such an ether is an unwarranted and unnecessary fantasy.

So long as we adhere to a unique identifiable ether we have the right to speak of absolute rest and absolute motion, for we then mean rest and motion relative to the ether. To discard this ether is to discard the notion of absolute motion and to regard motion as meaningless except in a relative sense, namely, the sense of motion of one material system relative to another.

The idea of absolute motion having been discarded, the choice of the system (frame of reference) in relation to which the motion of a body shall be described is arbitrary—it may be determined by convenience or simplicity for each particular case.

The striking nature of this conclusion is seen when the phenomenon of rotation is considered. From the standpoint of Relativity the phenomenon of a spinning top may be described with equal accuracy and completeness either by regarding the top as spinning relatively to the earth or by regarding the top as at rest and the earth—with the whole universe—as moving round the top. By either mode of description all the conditions, including of course the strains due to centrifugal force in the top, would duly appear.

Can the mind do without an ether?

It has been said that as the mind demands an ether there must be one. But what is meant by saying the mind demands an ether? Simply that without an ether to serve as transmitting medium the mind cannot "form a physical conception" of the transmission of radiation. But to "form a physical conception" of a phenomenon is found, on careful examination, to mean nothing more than this: to describe the phenomenon in terms of commonly accepted ideas—and this has meant, until quite recently, in terms of the ideas of Newtonian mechanics.

According to Relativity the notions of space and time underlying Newtonian mechanics are wrong, and physical conceptions based on these notions are wrong. It is advisable, therefore, to discourage the mind from demanding a Newtonian ether—whose only object is to make possible the formation of wrong conceptions.

How then is the Relativist to answer the question, "By what means is light transmitted if there is no medium for the transmission, and where or in what is the sun's radiation located during the time of transmission?" He will answer simply that such a question can be asked only by one who has failed to appreciate our position in regard to our world of phenomena. In no case whatever, not even in that of the simplest phenomenon, can the question "by what means?" or "why?" be answered. Explanations never explain, they merely describe, and all that scientists can do is to improve the old descriptions of phenomena or find new ones.

It is true that when one has become accustomed to descriptions of many varied phenomena all in terms of old and familiar conceptions, such descriptions begin to assume a friendly naturalness which leads us to think of them as explanations, but a little thought applied even to the most familiar of explanations discloses the fact that these are truly no more than descriptions.

The very idea of an explanation is beyond our understanding —just as is the idea of creation. All we can do is to take things as we find them—and among them ourselves, describing them.

Why not retain the ether?

It may be argued, however, that since whatever standpoint be taken, there remains inexplicable mystery at the root of things, we are hardly warranted in rejecting the old ether idea in favour of that of Relativity. The Newtonian ether idea does indeed involve, fundamentally, a few more factors that have to be swallowed as unanalysable pills, than does Relativity, but the one big pill of Relativity is harder to swallow than the many smaller pills.

The question thus raised goes to the root of the whole matter. The suggestion that it may be better to accept a large number of independent mysteries rather than have the mental labour of reducing the phenomena of nature to simpler terms, is one upon which we all do indeed act in daily life. We cultivate habits both of mind and of body which the crude and often involuntary process of trial and error teaches us are best adapted to the cycles of events through which life takes us, and we do this without worrying at all about what may be the meaning of these cycles of events, or why they happen. Or even when, in accustoming ourselves to some new cycle of events—as for instance the Fox-trot—we try to hasten the process by understanding the steps in the cycle, we feel that the explanation and the understanding and the manner of them are of no importance compared with the acquirement of the habit of going involuntarily, and without catastrophe, through the cycle.

If it were always merely a question of going through life with as little mental labour as possible the majority of us would certainly favour the old ether idea—simply committing to memory the long list of the inconsistencies involved—rather than that of Relativity, and on those numerous occasions when it is merely a question of saving labour we actually do this—or something cruder still.

Newton's laws and the notions of space and time which they involve will continue—probably for an indefinitely long time—to be used as the basis of all ordinary practical work. This means, in effect, that the old ether idea will continue to be used, for the ether was invented at the demand of the Newtonian conception of an absolute unique 3-dimensional space in reference to which the velocity of light, and, in addition, all accelerations, especially rotations, could be defined, "ether" being another name for this absolute space.

For the large majority of practical phenomena Newton's laws and the ether idea are adequate, while those for which they are inadequate seldom occur in ordinary work and those cases in which such unusual phenomena do occur may be treated, without inconvenience, as special.

The phenomena in reference to which the Newtonian ideas are inadequate are either those in which exceedingly great velocities occur—approaching the velocity of light—or those involving exceedingly precise calculations of mass and of gravitational force.

Relative velocities of such high order are impossible between masses of material large enough to be dealt with in engineering—because, first, no material is strong enough to impart to bodies big enough to be handled the acceleration needed to set up, within ordinary limits of space, such great velocities, and, secondly, a pellet of any known substance travelling relatively even to the atmosphere, with a velocity of such order would be volatilized by frictional heat. When, therefore, relative velocities of such high order do occur in ordinary engineering practice they occur only in the case of very small particles, as, for instance, the electrons in an electrical discharge. In such cases, just as in many cases of considerably smaller relative velocities (as, for instance, the relative velocities of particles of gas in a heat engine—velocities amounting to only a few hundred metres per second), it is practically never necessary to deal with individual particles, statistical methods being found adequate. In designing a steam engine the engineer considers not the motion of the individual particles of the steam but only that of masses of steam containing never less than many millions of millions of molecules. It is of no importance to the engineer to know whether or not the individual particles of steam are obeying Newton's laws of motion, he is satisfied to know that large masses of steam do so with sufficient accuracy.

Again, in all the cases of gravitational phenomena dealt with to-day or likely to have to be dealt with in the near future, Newton's law of gravitation—which is a special case of Einstein's more general law—and Newton's definition of mass as an absolutely invariable quantity, give results so exceedingly close to the truth that their error is hardly detectable.

To deduce, however, from such reasoning, that the Newtonian laws and the old ether idea ought to be retained and the Theory of Relativity discarded is to fail to appreciate the object of science, which is to find out experimental truth and so to be able to predict future events. It is only as a side issue that science saves labour, whether mental or physical—knowledge always does save labour.

The Theory of Relativity promises to provide a means of describing past events which is more consistent and involves fewer indefinable factors than the older theories—in fact it seems possible that when the theory in its "general" form has been further developed, all the phenomena of nature may be found subject to one inclusive law, namely, Einstein's law of gravitation.

We must take things as we find them

So long as phenomena present merely a succession of unrelated events they are without meaning, like a series of words in an unknown language, but just as the repetitions which occur in the written page may be made ultimately to disclose the meaning underlying the words and the laws of the language, so also in science the repetition of events and of ordered series of events leads to an understanding of nature and of nature's laws. Repetition of events leads to expectation of further repetition and upon such expectation are based our estimates of the probability of events which may occur in the future.

Thus our understanding of the meaning of things—for it is this which measures our ability to make good predictions of future events—depends on our reading of the records of past events. If the record presents them as a confusion of isolated unrelated facts, the reading of the record is impossible. Our aim is to produce a record presenting as few isolated factors as possible—an isolated factor being one which cannot be described in terms of factors already familiar.

However difficult of development the Theory of Relativity may be, however much mental labour may be required in comprehending it, however unfamiliar may be the terms in which it is expressible, the promise which this theory gives of providing a more perfect means of recording observations and describing events and so of enabling us more surely to predict future events, demands that the theory shall be fully investigated.

Is there an "ether" of any sort?

To many the Newtonian ether is an old familiar friend and, in spite of all its inconsistencies, a sense of loyalty makes them resent the suggestion that it ought to be discarded. Rather than discard they prefer to remodel it and are not discouraged when the remodelled thing is found to correspond in no point except its name, with the old friend.

The Theory of Relativity has no use for the Newtonian ether but it certainly leaves room for a properly chosen ether.

The world of Relativity is a continuum, in which are embedded mysterious singularities called, in ordinary language, particles, in the language of Relativity "world-lines," and a change occurring in any one little corner of it is accompanied by changes everywhere else. "Everywhere else" in the 4-dimensional space-time continuum means more than in space alone; it means at every other time as well as at every other place. Thus the hanging together of events in the continuum of Relativity is more simply stated than in Newtonian space and time, for the idea of one continuum is simpler than that of two independent ones.

The old ether represents the 3-dimensional space continuum of Newtonian science. If now the remodelled ether be taken to represent the 4-dimensional space-time continuum, it finds its place at once in Relativity.

In Newtonian space (the old ether) there are "points"; in 4-dimensional space-time there are "events" (an event being a point in space at an instant in time). The old ether is an "ether of points," the new one is an "ether of events."

We have objected to the old ether because it cannot be defined. Can the new ether be defined? Not completely. It is not, however, a final objection to a thing, that it cannot be defined. Fundamental factors never can be defined. This will be referred to later. Every proposition is based on earlier propositions and defined in terms of them; there must be at least one fundamental proposition—the first—which has no preceding ones to serve as its definition. In ordinary geometry the ultimate indefinable factors are the line and the point, and the basic propositions are the axioms and postulates.

With fundamental factors all that can be done is to name them and say they may be taken to mean anything one likes provided the chosen meaning is not inconsistent with what will be said about them. Thus the whole body of things said about these indefinables forms what may be called their "extended definitions."

Fundamentals never can be defined

The real trouble with the Newtonian ether is not that it is indefinable—in the ordinary sense of brief definition—but that the things said about it (its "extended definition") are inconsistent or contradictory to such a degree that no one has yet been able to form of it a conception (either physical or mathematical) which is not inconsistent with the things that have to be said about it.

In regard now to the "ether of events" we shall find that it is characterized by one property, namely that it serves as a "medium" for the space-time "ordering" of events. To define the "ether of events" it is necessary, first, to define "ordering," and to do this has been found impossible.

Relativity describes the world in terms of events and their space-time ordering. The description is possible only in one language, namely, the language of mathematics. Other languages (with the exception of the language of music, and this is certainly not adapted to unambiguous expression) are "picture" languages, properly adapted only to those modes of conception known as 3-dimensional space and 1-dimensional time. These languages can describe the 4-dimensional world of Relativity only by analogy. Mathematics, on the other hand (the most logical of all languages) is not a "picture" language; it can express and describe events in a world of many dimensions just as completely and directly as in a world of three. Often a "physical conception" of its statements is found impossible, but we shall see that the possibility of such a conception is of no essential importance in a scientific description.

In the language of mathematics an unlimited variety of descriptions of the world is possible, but in all of them there is found one unchanging factor, namely, the space-time "ordering" of events, and it is found that to leave "ordering" without definition, but to say of it that it may be taken to mean anything one likes, provided the chosen meaning is not inconsistent with the things said about it, leads to no ambiguities or contradictions.

Is there absolute motion in the ether of events?

Does the recognition of an ether of events re-instate the idea of absolute motion? The answer is no, because "motion" is meaningless in respect to the ether of events.

The Newtonian world is a 3-dimensional space continuum (the "ether") with particles *in motion* in it—time being necessary to this motion. The world of Relativity is a 4-dimensional space-time continuum (the "ether of events") with "world-lines" *embedded* in it: not world-lines *in motion* in it, for the idea of motion belongs to the Newtonian space.

The 4-dimensional world is a "static" world, with the events of the past, the present and the future all definitely located in it. To the universal observer a human being is a sheaf of world-lines.

World-lines may be pictured in reference to the continuum formed by 2-dimensional space and time. We take, for instance, the 2-dimensional space defined by the surface of this page. The full-stop just written represents a particle at rest in this space, and in order to show the particle's progress in time we imagine that the full-stop moves—or rather "exists"—downwards through the pages of the book, each succeeding page, while having the same space significance as this page, representing a later instant in time. The path of the full-stop is thus thought of as a line drawn perpendicularly downwards from this page. This line is a "world-line," and represents the particle completely, that is to say in its continuous being.

If the chosen particle were not the full-stop but were the tip of my pen as I write this page, then the world-line considered would be a strangely bent and twisted line meandering downwards through the pages. As a particle's progress through time is essentially continuous—for it cannot halt without spending time—the words must be supposed written without raising the pen from the paper.

The more quickly my pen moves from left to right the more oblique to the time direction (which is perpendicular to the paper) does the world-line become.

Thus there is no velocity or motion, as such, in the space-time world, but the slope of its world-lines is the counterpart of what we human beings, with our 3-dimensional space, call "motion."

In the "ether of events" there is no motion either relative or absolute, but there is absolute world-line structure, and this is the absolute reality underlying "motion."

CHAPTER I

The search after truth

LOOKING for truth, like hunting the thimble, is a process of trial and error. A phenomenon is observed, a guess is made as to why it happens and the guess is then tried out. If the guess does not work it is discarded and another is tried. The failure of a guess is not useless, for it gives a lead as to what sort of guesses to avoid in the future, so that, in time, by the slow process of eliminating what is not the truth scientists find out what is the truth.

A guess should be expressed in words whose meaning is unambiguous, otherwise its working may be mistaken for not working, and, since unambiguous definition of the differences between two phenomena can be made only in terms of measurements, the factors expressing a guess should be capable of quantitative measurement.

The famous guess, "Nature abhors a vacuum," has had to be discarded. It satisfied the ordinary man and was obviously difficult to refute, but the vacuum pump showed conclusively that nature's abhorrence was a thing which could be overcome, and that one never could be sure to what extent the abhorrence would work, nor indeed, in some cases, whether it would work at all. But the chief weakness of this guess lay in the vagueness of the term "to abhor" and its unsuitability to unambiguous discussion.

Scientists call a guess a hypothesis; if it fails to work it is a false hypothesis; if it sometimes works it becomes a theory; if it always works it becomes a law of nature and so a part of scientific truth.

A guess which sometimes works but not always, may be called either a theory or a partial truth; to be an unqualified truth it must work always. Obviously then, since experience is limited and trials can never be exhaustive, no truth can be called, legitimately, more than a partial or tentative truth.

In defining truth as that which works (a definition often called pragmatic or practical, to distinguish it from the essentially ambiguous definitions which find favour with most men, and whose chief characteristic is that they engender strife) it is important to say who shall be the judge of the working. Only recently has the importance of this come to be realized. It used to be thought that if two different observers gave two different descriptions of an event, the differences either were such as could be eliminated definitely by correcting for the different conditions under which the observers worked, or were due to actual errors of observation. It is now known that cases exist in which two different observers, adhering to all the rules which—until the advent of modern Relativity—had been considered sound, must give inconsistent descriptions of one and the same event, and that no criterion exists as to which of the two descriptions is the more worthy of respect.

Such a state of affairs is not easy to grasp, for in thinking of any physical fact to be observed we naturally attribute inconsistencies to errors of measurement. For instance: Two distant shots are heard. One observer makes measurements which prove to his own satisfaction that the guns were fired simultaneously; another finds that there must have been a time interval of one second. Which result is true? If the relative positions of the guns and the observers were known, and the time interval between the sounds as heard by each observer, then a knowledge of the velocity of sound should enable the correct interval between the firing of the guns to be deduced—provided there was no wind. If there was a wind then its velocity and direction would have to be taken into account by each observer. Thus some third observer, who had taken all these factors into account, should be able to decide between the two results. At any rate there could be no question but that inconsistent results meant error of observation.

Some light may be thrown upon the possibility of cases in which observers may get inconsistent results although they all make their measurements with accuracy, by considering, as an illustration, how the question "which result is true?" would stand in the practical case just instanced, if it had been impossible to determine the direction or velocity of the wind.

We simply should have been unable to say which or whether either of the two results might have been correct.

But, it will be said, such an attitude would be unwarranted because we have other means besides that of sound for the determination of the time interval. This is so, but, for the purpose of our illustration, we must assume that the noise of the shots was the only evidence of the firing of the guns and provided the sole means of judging of the time interval between them. We then have a case analogous to what is an actual case in optics when flashes of light are emitted from two separate sources.

As an aid to conception it has been usual to postulate a medium—the ether—whose function is to transmit light (as well as every other form of radiation), and just as air transmits sound with a velocity of some 1100 ft. per second so it is supposed that the ether transmits radiation with a velocity of 186,000 miles (= 3×10^{10} cms.) per second. The ether is, in many respects, a very useful invention, but it has proved itself intangible. No observations of any kind can be made upon an identifiable piece of the ether and it is unknown whether anything of the nature of an ether wind ever occurs. Hence, in the case of light, and of all radiation, scientists are faced with just the sort of difficulty which has been mentioned in connection with sound transmitted by a wind of unknowable direction and velocity. They have to admit, therefore, that the time interval between distant flashes of light is a practically unknowable quantity. They go further—if they are Relativists—and say it has no right to be called a real quantity at all.

Time intervals are unknowable

Different observers obtain different values for the time interval between two given flashes and not one of them can substantiate a better claim than any other to be right. This simply means that time intervals are unmeasurable things; hence, to the Relativist—who says that anything which is real must be measurable—time intervals are not real things belonging to the event observed, but are quantities imputed to the event by the observer.

The idea to which the mind persistently clings that, no matter what experiments may say or scientists may think, there must indeed be one and only one "true" time interval between the flashes, is partly the expression of one's natural ingrained solipsism—and therefore is irrefutably true, or, in other words, is a matter of taste—and partly the outcome of the mind's dogmatic belief in a real identifiable ether. We are unable to find the velocity of the ether wind, but, according to this dogma the ether either is stagnant or does possess some definite velocity. Our ignorance of this velocity entails ignorance of time intervals but these intervals are none the less real.

Anyone who insists on holding this dogma must be regarded as unduly credulous unless he is able to prove the existence of the ether. The first step towards such proof must be to define "ether," no consistent definition having yet been found for it. The ether was invented as a medium by which radiation might be transmitted. Experiments, both optical and electrical, have shown, however, that either the ether possesses the contradictory property of being at rest relatively to all observers no matter how these observers may themselves be in motion relatively to one another, or it possesses the property of causing bodies to contract and clocks to go slowly when they move relatively to it. These two supposed effects are, it has to be admitted, undetectable, because instruments used in the attempt to detect them must be correspondingly affected. Thus the curious property which has to be assigned to the ether to save it from self-contradiction proves to be a property which has to be taken on faith—it is a property which the ether must be supposed to possess in order that the ether itself may be supposed to exist! And this Newtonian ether is supposed to exist, for no practical purpose at all (since Relativity shows its existence to be a hypothesis unnecessary to the description of any known phenomena) but merely to support the pre-conceived and unwarranted dogma that the Newtonian notions of space and time are "true."

According to Relativity the relative magnitudes of the different time intervals assigned to events by different observers will depend upon the velocities of the observers relatively to the things observed. Thus it becomes important, in discussing a phenomenon, in order to determine whether a given proposition works or not, to know the motion relative to the events, of the observer who records the phenomenon.

What is "reality"?

Einstein has attacked the problem of such cases in general and, in order to avoid the ambiguity which, as has just been explained, is inseparable from descriptions of phenomena in terms of space and time measurements, has sought to express the "reality" underlying such measurements.

A thing can hardly be "real"—or at any rate it can hardly substantiate a claim to reality—unless it can be defined, and a definition is practically useless unless it is unambiguous. Freedom from verbal ambiguity is not sufficient, but in order to ensure such freedom from ambiguity as shall make discussion possible, it is essential that a definition shall be understood in a similar or equivalent manner by all men. In discussing, for instance, the question, what is the real value of gold, it is essential that all parties to the discussion shall accept the same definition of "value" and of "gold." Gold has a definite "real" value only to those who have a similar understanding of value, and since a practical agreement on this matter could hardly be reached except in a limited community, it is usual to regard "value" as a relative quantity, not a real quantity at all.

Definition of "reality"

For reasons of this sort scientists have taken a definition of reality which runs, "that and that alone is real which is perceived by all observers." Here "perceived" means similarly perceived, for if two perceivings differed, the factors by which they differed would be perceived, not by both observers, but only by the one whose perceiving contained them.

This definition of reality can be put into other forms, for instance, since a thing is not scientifically perceived unless the perception can be defined, and definition, to be unambiguous must be quantitative, one form the definition of scientific reality takes is: "That which can be measured is real."

The space-time continuum

As applied to such questions as whether a theory is true it reads: "That and that only is true which works to the satisfaction of all observers." Einstein has accepted this definition and has found it necessary to reject the hitherto accepted notions of space and time and to replace them by the conception of a 4-dimensional space-time continuum.

When observers, using the Newtonian conceptions and equipped with similar rods and clocks, set out to measure, say, first the length of a rod of steel and then the time interval between two events, it has been customary, hitherto, to assume that the results must all agree or be capable of being brought into agreement by properly applied corrections. It was assumed that in all cases such possible differences as those of position or of motion of the things observed relative to the observer could be allowed for, and if two observers obtained different results one or both must be in error. There could be no possible doubt but that more careful observation could discover what was indeed the "true" length of the rod and the "true" time interval between the events.

For all ordinary practical purposes such as selling tape and catching trains—or even making lenses, accurate to the one millionth part of an inch, or chronometers correct to a second in a week—the old fashioned assumptions just spoken of are satisfactory, for although owing to the possible existence of an ether wind the lenses may have suffered enormous distortion and the clocks may be going slowly, these conditions are of no importance because the users of the lenses and clocks and all their standards will have suffered compensating changes, while the relative velocities likely to exist in practice between the observer and the thing being observed, are too small to cause detectable differences in the Fitzgerald effects.

In modern scientific work, however, such as the measurement of the "diameter" of a swiftly moving electron, or of the time taken for light to travel across a rigid material system

Chief [illegible]

in rapid motion, the old assumptions have led to inconsistent results, not one of which, however, can be proved more "true" than another.

Einstein's guess

The recognition of the existence of such inconsistencies led Einstein to guess that there must be something wrong with the old notions about space and time. All the observers worked accurately, all had similar and accurate measuring devices, yet they found different values for the length of one and the same rigid body and different values for the time occupied by one and the same event: the inconsistencies must surely be attributable to something in the notions these observers had as to the quantities—space and time—which they were observing. It must be that the "length" of a rigid rod was something attributed to the rod by the observer, but not "really" belonging to the rod, and the time interval between two events something attributed to the passage of the events and not really belonging to the events.

So novel is this point of view and so contrary to common-sense that, in spite of the splendid success of the theory which Einstein has based upon it, this theory is being accepted with reluctance. The difficulty of accepting the theory may be increased so as to become almost insuperable, by being left nebulous. Some indefinable, but, we feel sure, highly respectable intuition tells us that space and time must not be juggled with, and that any explanation which succeeds only through sacrifice of the most real things of life—namely, our intuitive notions—must prove ultimately to be wrong.

In order to be able to meet such difficulties—which, usually, are not serious difficulties but rather are wordy confusions—it is well to begin by defining them.

The difficulties will be found to arise out of the definition, already mentioned, which Einstein takes of truth, namely, "Truth is that which works to the satisfaction of all observers." In its equivalent form, "Those things are real which appear similarly to all observers," this definition is obviously in line with ordinary ideas; for instance, a mountain is regarded as real; there is no ambiguity about a mountain, people do not quarrel about its existence. A "ghost," on the other hand, is unreal, it is seldom observed by more than a few people and the observations made by these few never agree; moreover, in many cases, when reasonable correction has been made for the condition of the observers at the time, the "ghost" can be explained away.

Ghosts

A form of the same definition often favoured reads: "That which can be measured is real." Here the word "measured" is used instead of "observed" and with a wider meaning, namely, the meaning of unambiguously measured in quantities which all observers would find the same.

It is difficult to invent a more useful definition of truth than Einstein's but easy to invent a more personal one, and it is because this definition lacks respect for individual personal points of view that we find the theory difficult which has been based upon the definition.

We are all solipsists at heart, and though for peace and practical convenience we are willing to compromise, we know that in truth the only realities are our own personal feelings. Hitherto the compromises we have had to make have seemed unimportant, but the demand now made that we shall not merely compromise about but be prepared to reject our fundamental notions of space and time, is difficult to meet with equanimity.

We might manage to bear a mere adjustment of times and lengths to suit different observers, but it seems too much that we should be asked to concede ignorance of the time order of events or that time order is a personal matter,—the suggestion is simply immoral, for if my future, from any conceivable observer's point of view, can precede my present then my future must have been predetermined and I am not to blame for anything I do.

But thus to allow the impersonal standpoint of the Theory of Relativity to lead us into difficulty is to misunderstand the object and meaning of the theory.

Scientific guessing has as its object the prediction of future events. A good guess takes into account all past events, sees what events and series of events have been repeated in the past and deduces the probability of further repetition in the future. An effective survey of the past is possible only if the record of past events is completely intelligible and "true."

Einstein proposes that since experience shows that records made by individual observers in terms of the old conception of space and time involve irreconcilable inconsistencies, it is impracticable to regard such records as true until the "reality" underlying them—which by definition must appear the same to all observers—has been extracted and expressed.

The universal observer

To treat in this way the records made by individual observers, is to put them into the language of what may be called the "universal observer."

Individual observers are biassed by their states of motion relative to the things observed. To find out "truth" it is necessary to discover what are the factors in those varying records which are common to all conceivable observers, or what transformation of the contents of the various records will produce such common factors.

The "universal observer" picks out automatically these desired factors. He has eyes and fingers in all the corners of the universe and these eyes and fingers can assume every possible state of motion relatively to the things they are observing. To him the world is very different from that of any individual human observer's experience. It is a 4-dimensional world in which space and time play similar rôles.

It is foolish because irrelevant to quarrel with the universal observer on the ground that his world is fundamentally different from one's own. The 4-dimensional world has been invented with the sole object of systematizing records of past and passing events so that they may be used effectively in the prediction of future events; it is the "true" world because it has been made consistent, at every step in its development, with the arbitrarily chosen definition of "truth."

It is legitimate to quarrel with the universal observer on the ground that his world is based on a definition of truth of which, personally, one cannot approve, but it is no easy matter to invent a better definition of truth.

What is truth?

Although it is difficult to invent a definition of truth which, for the purpose of science, is better than the pragmatic one just discussed, some people feel that the definition is artificial; nor is the feeling removed by reiteration of proofs that it is without scientific basis.

Pragmatic truth

In order to brand the obnoxious definition it has been given the awkward name "pragmatic." The plain word for "pragmatic" is "practical," and to object to the definition because it is pragmatic is to object to it because it is practical—that is to say derived from experience. Why then does anyone object to a definition based on experience?

Some have argued that knowledge of truth comes from something deeper than experience—nor are they discouraged by the fact that experience shows them to be wrong; experience being, for them, no criterion of right and wrong. Such a point of view leads a man to seek truth by shutting himself in a cloister.

Practically speaking it is nonsense to suggest any basis for knowledge other than experience—as of course it must be, since sense is experience—but before dismissing the matter in this way we ought to make sure that no reasonable point of view has been overlooked.

Do the anti-pragmatists object merely to the special practical meaning attached by science to the expression "that which works"?

It is only recently that the necessity has been recognized of adding the qualification "to the satisfaction of all observers": previously it had been assumed that such a qualification was unnecessary, the possibility being then unknown, that different yet equally legitimate opinions could be held as to whether a proposition "works" or not. And possibly the ground of the anti-pragmatist's objection lies in this qualification.

The suggestion then is that the anti-pragmatist takes, in effect, a definition of truth which reads, "Truth is that which works to the satisfaction of myself." If this is indeed the position, it becomes clear why the anti-pragmatist objects to a definition of truth which makes of the world a 4-dimensional continuum in which space and time play interchangeable rôles.

As a mere individual I *know* that the things I mean by space and time, though, possibly, I cannot define them, scientifically, are *real* and are essentially separate and distinct, and that time order as perceived by me is absolute. I know this because I feel it, and, to me, my feelings are absolute reality.

Solipsism

Thus defined the position of the anti-pragmatist is that of the solipsist, and it is impregnable. It may be that all men are solipsists at heart; when, however, they play the part of scientists they become pragmatists—their object being the crude and practical one of predicting future events.

CHAPTER II

The problems of life

To know the meaning of things is to foresee their consequences, and if we knew the meaning of all things we could predict the future with certainty and so, presumably, we could better ensure our own future happiness—at any rate such a supposition seems to be supported by two striking facts, namely, that the desire for future happiness is the chief incentive of men's lives, and that all men (and especially scientists, philosophers and theologians) consider it highly important to find out the meaning of things.

Not that they can hope to do this in an absolute sense. It may be that all the things of life are explicable in terms of a few simple factors, but since, after all, we ourselves form a part of our world it is unreasonable to hope that the ultimate factors can themselves be explained by us.

Matter and electrons

Matter, for instance, has long been thought of as built up of molecules, which themselves are built up of atoms. Atoms were for some long time regarded as the ultimate components of matter: they were called atoms to indicate that they could not be cut up into parts, could not be analysed or explained. It is now known that atoms have a structure and can themselves be cut up into much smaller things, namely, electrons. Electrons play, for the time being, the part of ultimate, indivisible, inexplicable components of matter. Should it be found that electrons have, themselves, a structure then the elements of this structure will displace the electron from its position as an ultimate thing.

Every explanation must have a basis; starting from nothing we can explain nothing; every argument must have premises and if the argument is fundamental the premises themselves must be beyond argument. Such premises are called assumptions.

Indefinable fundamentals

In Euclid—to take an example from geometry—the fundamental assumptions are the axioms, definitions and postulates, and as the Euclidean is not the most fundamental geometry, it is possible to argue about the meaning and the truth of its basic assumptions—indeed it is possible to replace them by others, quite different, upon which equally good systems of geometry can be built. Moreover, in Euclid, the inexplicable nature of the fundamental elements (points and lines) is obscured. The point and the line are supposed to be "defined," and the student feels that they have thus been explained. Modern geometry recognizes that such definitions are a mere quibble, and instead of giving a definition of a point it says simply: The meaning of the word "point" cannot be defined; the word may be taken to mean anything the reader likes, provided only that the meaning assigned shall be not inconsistent with the things that geometry says about the "point."

Thus the whole system of geometry may be said to constitute a sort of "extended definition" of the point and the line.

Similarly in explaining the phenomena met with in life the best we ever may hope to do is to get down to ultimate factors which, themselves, defy explanation. Of such factors we may say: They may be taken to mean anything one likes, provided only that the meaning assigned shall be not inconsistent with what the scientists and philosophers and theologians say about them.

The worthiness of these factors to be called "real" is known by the possibility of assigning to them a meaning which is indeed not inconsistent with the things said. Up to the present neither physical science nor philosophy nor theology has succeeded in naming fundamental factors to which the task of assigning such a meaning has proved a possible one. Physical science has been, however, much less unsuccessful than either philosophy or theology, and the coming of Einstein's General Theory promises to increase immeasurably the advantage thus possessed by physical science.

The meaning of "consistent"

In the search after the truth underlying things it is of fundamental importance to assign a meaning to the words "consistent" and "inconsistent." We are allowed to attach to the indefinables any meaning "not inconsistent" with the things said about them. What, then, is the criterion of consistency? Two statements are not inconsistent if they are not contradictory, but what is "contradictory"? It is because of the great difficulty of defining such terms as these that communities of comparatively sane men are able, with sincerity, to "believe in" an unlimited and weird variety of "isms."

Physical science is fortunate in having restricted the issue in regard to "truth," within reasonable limits. The restriction is arbitrary but without it science would be useless. In physical science truth is that which works, and reality is expressible in measure results. We shall see, later, that every "measurement" consists in observation of a space-time coincidence—a material encounter or its assumed equivalent. In science the absolute realities are space-time coincidences and their space-time "ordering," and inconsistency consists in stating that a space-time coincidence both is and at the same time is not, as, for instance, in saying that a pair of particles collided and at the same time did not collide—that a "hit" was at the same time a "miss." Another inconsistency would consist in saying that the "order" of three coincidences A, B and C in space-time was both ABC and ACB. Space-time ordering is not space ordering nor time ordering. There is no scientific inconsistency in saying that the time order was ABC to one observer and ACB to another or that the space order was different to different observers, for time relations and space relations are relations imputed to events by the observer, they do not belong to the events.

Space-time order is defined in science as absolute because it is an invariant in the description of phenomena made by all conceivable observers. It is the "order" of occurrence of crossings of "world-lines."

The old ether is having to be discarded because, as it possesses no identifiable parts, space-time coincidences (encounters) are meaningless in regard to it. No measurements can be made upon the ether and no measure results can describe it—it is "unreal."

Truth which is deeper than that of science

An advantage has been claimed for physical science over philosophy and theology, but it should be remarked that the advantage is an artificial one.

Physical science has chosen an arbitrary definition of "truth" which simplifies the problem of finding out truth. The choice is warranted only because it enables science to be useful, and only to the extent that practical usefulness is a desirable object.

The position is humanly untenable that practical usefulness is the greatest object in life.

There are deeper truths than those of physical science and it is with these, as well as those of science, that philosophy and theology concern themselves. These truths relate to individual human feeling and experience, and the "universal observer" is blind to them.

It is hardly to be wondered at that, discouraged by the great difficulty of finding a fundamental definition of truth capable of being consistently held, philosophers and theologians are constantly being tempted into arbitrary (dogmatic) definitions. It is only in this way that the problem of building up a "system" seems practicable.

Scientific knowledge is of relations

Geometry takes points and lines and describes their possible relations, which are relations of "space order." Science takes the most fundamental factors it can find and describes the world of universal experience in terms of the relations of these factors. Both geometry and science are thus systems of relativity. They deal with the relations of ultimate undefinable things, which themselves cannot be explained. These ultimate things are often called "absolute." In science the things which to-day play the part of absolute elements may to-morrow be analysed—that is to say split up into still more fundamental elements—and experience suggests that this process may never end.

The nature of "explanations"

Explanations can be nothing more than descriptions but descriptions are only worthy to be called explanations when they simplify.

It is not easy to define simplification. Brevity is said to be the soul of wit, but it is not always that an abbreviated description is a more explanatory or a simpler one. The meaning of simplification may be arrived at, perhaps, by the consideration of some instances:

(1) The fraction 2256/4512 may be simplified into 1/2. Here the simplification consists in reducing the number of symbols. The object of the fraction is to express a relation, without regard to either the nature or the quantity of the things related. The relation might be expressed by writing (2256 apples/4512 apples) but this form of the expression introduces, unnecessarily, the hypothesis that apples are being thought of. It is simpler to write 2256/4512, an expression which, by ordinary convention, means the relation of 2256 things (of whatever sort or mixture of sorts) to 4512 things, that is to say a mere numerical relation. However, the form 2256/4512 introduces the unnecessary hypothesis that the ratio of those two particular numerals is being thought of, and the reader is inclined to suppose that some particular case is being referred to, as for instance the number of men and the number of women in a certain town. The alternative form, 1/2, defines completely the numerical ratio intended and, by convention in regard to all fractions when in their lowest forms, it should not be taken to mean more than a ratio.

This example suggests that simplification consists in the removal of what is unnecessary, whether this consist of symbols, hypotheses or suggestions.

(2) The algebraic expression $(a + b)^2$ may be written $(a^2 + 2ab + b^2)$. The form $(a + b)^2$ looks simpler yet the difference is only formal. Both forms are complete, and neither contains anything unnecessary, that is to say anything whose removal would leave the meaning of the expression unchanged.

In one sense $(a^2 + 2ab + b^2)$ may be considered as explained by $(a + b)^2$, for this says—taking, for instance, the special case in which $(a + b)$ is a straight line divided into the segments a and b—that the sum of the squares on the two component segments of a line, plus twice the rectangle contained by the segments is equal to the square on the whole line. In fact the statement $(a + b)^2 = (a^2 + 2ab + b^2)$ is Euclid, Book II, Proposition 12. If then we consider that $(a + b)^2$ is in fact no simpler than $(a^2 + 2ab + b^2)$, or, conversely, that the latter is no more complex (or in need of explanation) than the former, we shall have to say, consistently with this, that none of the propositions of geometry, however advanced it may be, is, in fact, more complex than those which precede it and form its basis, or, ultimately, than the axioms, definitions and postulates on which the whole system of geometry is based.

The same principle may be extended to the whole of mathematics, since in none of the steps in a mathematical argument is anything contained which was not arbitrarily put in at the outset. However, the conclusion thus reached, that no mathematical process can be called a simplification or an explanation, is evidently contrary to generally accepted ideas. We must allow that mere formal processes often do simplify or explain.

Brevity

This leads to the thought that one of the functions of an explanation is to make a matter easy of understanding to the ordinary human mind. Brevity need then be no criterion of the value of an explanation; in fact rather than the removal of what in the strict sense are unnecessary factors, the insertion of factors which, later in the argument, will have to be removed, is often desirable, if the explanation is to be a really good one. Thus, in geometry, lines often are introduced "for construction." Such factors may be compared with the temporary scaffolding used by builders to simplify the process of building.

Sometimes the "unnecessary" factors thus introduced for the simplification of an explanation may be unwarranted hypotheses. There are, for instance, some problems in the science of "heat," whose explanation may be simplified by the hypothesis of a heat substance, or in electricity in which explanations in terms of the "water analogy" are helpful.

(3) As a third example we may take the famous law by which Mendelejeff brought order into what had appeared to be random differences in the properties of the chemical elements. Mendelejeff showed that when the elements were arranged in

the order of their atomic weights, at intervals along a wave-shaped curve of simple geometrical form, then those elements which occupied similar positions in relation to the crests of the wave would have similar chemical properties. The elements could thus be grouped into families and the occurrence of obvious gaps in some of the families led to the prediction of new elements which afterwards were actually discovered. Thus the application of a geometrical construction readily grasped by the mind simplified the understanding of chemistry.

This example suggests that the human mind regards a matter as having been simplified and, to some extent, explained, by a process which does no more than assist in the memorising of events or elements in the matter concerned

(4) There are two well-known modes in which the relative motions of the members of the solar system may be "explained"; one, the older geo-centric mode, in which the earth is regarded as the fixed body of reference; the other, introduced by Copernicus, in which the sun is regarded as the fixed body of reference. Both are legitimate; both give a correct description of the motions, but the Copernican is by far the simpler. Around a fixed earth it is true that the sun and the moon describe simple, almost circular paths, but the paths of the sun's planets and of their satellites are complex curly lines difficult for the mind to grasp and awkward to deal with in calculation, while on the other hand around a fixed sun all the more important paths are almost circular.

From the point of view of this example we may conclude that the virtue of an explanation depends largely upon its adaptation to the limited powers of conception of the human mind.

Consideration leads then to the following suggestions:

An explanation can be no more than a description, and the following characteristics are desirable:

(1) It should describe in terms of fundamental factors, the ideal fundamental factor—called "absolute"—being so fundamental that no description of it is possible. An absolute factor

may be given a name and it may be conceived in any desired way, provided only that the conception is not inconsistent with what is said about the absolute factor, in the explanation.

(2) The number of fundamental factors should be a minimum.

(3) The number of symbols and of hypotheses used in the explanation should be a minimum.

(4) The explanation must be easy of understanding and to this end should employ only such conceptions as are in common use.

It has been shown that (4) often is inconsistent with (1), (2) and (3); indeed (4) makes the value of an explanation an entirely relative matter; an explanation which for instance might be good from the point of view of a trained thinker would be worthless from that of an uneducated man.

This difficulty disappears if explanations are classified according to their objects.

The objects of explanations

Since there is pleasure in the exercise of the mental faculties one of the objects of some explanations is to afford such exercise. In these cases brevity and the avoidance of unnecessary hypotheses are unimportant. These might be called "tutorial explanations." They have, however, a value additional to that of giving pleasure, namely, that of training the mind to discovery. Knowledge comes through explanations. As soon as a phenomenon, first met with, perhaps, in nature, has been explained, it often becomes possible to reproduce the phenomenon artificially, to some immediately useful end. We may name, therefore, as the second and more important object of explanations, the increased knowledge which they bring. An explanation which has this as its sole object need not be easy of conception: it need not, indeed, be possible of conception. All that is essential is that the explanation shall be logically correct and actually true.

We must examine the statement that an explanation need not involve conceptions which are mentally possible.

Familiar conceptions

Examples will be taken first of explanations in terms of conceptions which are not only possible but familiar.

(1) The complex phenomenon of the motion of waves upon water may be explained by aid of the familiar conception of the motion of a pendulum. Of the latter everyone has a clear physical conception, and when proof has been given that the law underlying the motion, namely, the law expressed for the pendulum case by the familiar equation ($T = 2\pi\sqrt{l/g}$), is applicable also to the elements of a wave, it is felt that the phenomenon of wave motion has been well explained.

(2) The behaviour of gases when heated, consisting but a few years ago in phenomena which had to be observed, recorded and wondered at, became, through the invention of the mode of explanation known as the Kinetic Theory of Gases, a comparatively simple matter, of which the mind could form a physical conception. By this theory a gas is regarded as made up of a large number of very small, perfectly elastic, spherically shaped particles, each possessing a fixed mass and an average energy of motion proportional to the absolute temperature of the gas. The application of Newton's laws of motion to these particles suffices to explain the behaviour of all gases when heated or cooled or compressed.

Examples of explanations which, though answering the test of usefulness, are, nevertheless, impossible of conception, are no less numerous than the others but custom has made us blind to their difficulties:

(1) The most outstanding case is that of the force of gravitation. Newton's law of gravitation has explained gravitational action so well as to make possible the prediction with a remarkable degree of accuracy, not only of the phenomena of falling bodies near the earth's surface, but also of those of planetary motion. Eclipses are predicted with precision and new stars have been discovered. Nevertheless, the explanation is one which involves a process impossible of conception—namely, that of "action at a distance." We can conceive of a force acting through a material medium, but Newton's explanation

of the force of gravitation demands no medium; two masses attract each other with a force proportional inversely to the square of their distance apart even though the masses be separated by empty space, or, if there happen to be an intervening medium, Newton's explanation allows this to interfere in no way with the transmission of the force.

Action at a distance

Human minds—or at any rate many very reputable human minds—revolt against the notion of such "action at a distance," regarding it as absurd. Nevertheless, Newton's explanation answers the test of usefulness better than any other explanation of any physical phenomenon ever has done. As an explanation Newton's law of gravitation stands very high; the fact that it has now been shown to be only a special case of the more complete law invented by Einstein, detracts in no way from its practical usefulness.

(2) Electrical science has but few explanations which are possible of conception. Electrical and magnetic forces appear to provide examples of action at a distance but they differ from gravitational forces in this, that the presence of an intervening medium does affect the forces. It was because of the mind's repugnance against action at a distance,—the transmission of force and of energy through space containing no medium,—that the Newtonian ether was invented. Given an all-pervading medium, mental conceptions could be formed of electrical action. Unfortunately however all efforts to substantiate the reality of the ether have failed. Nevertheless, explanations of electrical phenomena, whether based on the notion of action at a distance or on that of action through the ether—and in either case impossible of conception—do answer, with remarkable success, the test of usefulness, and by their aid advances in knowledge and in practice have been made, comparable with, and perhaps greater than those made in mechanical science—all of whose explanations are said to be eminently possible of conception.

Possibility of mental conception no criterion of the value of an explanation

The conclusion to which these examples point, namely, that possibility of mental conception is no criterion of the value of an explanation, is so striking that it may be well to examine more carefully into the meaning of the expression "possibility of conception."

It has been tacitly assumed that the reader will agree, first, that ordinary mechanical phenomena are possible of conception and secondly that action at a distance is impossible of conception.

A little thought will show that the example of the pendulum as a simple conception was hardly consistent with the idea of the inconceivability of action at a distance, for the action of the pendulum depends on gravitation. It would have been better to have chosen, instead of the pendulum, a loaded spring, as, for instance, the prong of a tuning fork. Here also the motion is typical of wave motion. We thus avoid ordinary gravitation, replacing it by the elastic forces of a spring—which then we must assume to be readily capable of conception. But the action of a spring is only a special case of the property of the elasticity of bodies in general. We bounce a ball by throwing it horizontally at a vertical wall—the action is familiar, and easy of conception. The elastic forces called into play suffice to reverse the ball's motion. There seems to be here no need of any such conception as that of action at a distance. The close contact of impact and the resulting compression of the ball, followed by the elastic restoration of shape are regarded, ordinarily, as processes which commend themselves to common-sense. There is nothing, here, of the absurdity of gravitational action, in which one body moves towards another with increasing velocity although no conceivable means are available by which even a message may have passed between the two bodies to tell them they ought to approach each other.

Possibly a man like the famous Home, who is said to have moved distant pieces of furniture by mere mental effort, may have regarded action at a distance as a simple conception. We ordinary mortals can move things only by pushing or pulling and these actions demand material contact. In fact it seems as though one of the simplest of physical conceptions is that of contact and another that of elastic action.

What is "contact"?

But what really is "contact"? Is there indeed no distance between two bodies which touch? An instance of very close contact is that of two pieces of iron welded together. Contact is here perfect; and, in general, we can think of no contact more perfect than that between adjacent particles of a piece of solid matter; there the particles actually touch one another. How do we then adjust our minds to the conception that even a block of iron may be made smaller by compression?

If the particles of the iron are already in contact they cannot become more closely packed—unless perhaps the original packing represented a porous mass; a conception which, though applicable to solids, is surely difficult of application to liquids, and these also are compressible.

If we assume that the packing of the molecules of a solid is such as to leave interspaces—like the holes in a sponge—we have to ask why, after moderate compression, the solid resumes its original shape and volume? The reply is that the material is elastic. We can hardly say that the empty pores do themselves exert forces; the forces must be due to the walls of these pores, and these walls at any rate must consist of absolutely close-packed particles. The elasticity of a mass of close-packed particles then has to be attributed to the particles themselves. If these particles are the molecules we know that they have a structure; they are built up of atoms and built so loosely that the atoms can move about within the molecules. We might attribute elasticity to change in the closeness of the packing of the atoms, but to the atoms themselves must then be attributed the real elasticity.

Compression of an atom is conceivable because an atom is itself a very loosely packed system of electrons. Thus, ultimately, we have to conceive of an electron as itself elastic. It must be capable of being compressed and afterwards of resuming its original shape and volume. If the electron were known to have a structure we could continue this line of reasoning a step farther and attribute the elasticity to the component elements of the electron—and so on *ad infinitum*, the mysterious property of elasticity being thus attributed to the last member of an unending series of ever diminishing particles! But what has then become of the clearness and simplicity of our conception of contact and of elasticity?

Actually, at the present stage of scientific knowledge, the electron has to be regarded as structureless and so, indivisible, and we cannot shift the difficulty beyond the electron. The electron has to be thought of as having no component parts. How can we then conceive of its possession of elasticity? If it has no parts the idea that it can suffer change in volume is absurd, for, if we suppose it to have suffered a compression of, say, 1 % of its original volume, that little element of volume must have been occupied, previously, by a part of the electron—a notion which is inconsistent with the assumption that the electron has no component parts! Neither can we conceive of an electron suffering change of shape, since change of shape means re-arrangement of component parts.

In fact the ultimate particle of matter presents great difficulties: it need not be the electron—probably is not—but the atomic notion of the constitution of matter does surely demand an ultimate particle, and such reasoning as has been suggested shows that to this ultimate particle no properties of any sort—not even magnitude—can be assigned.

The alternative of pushing the responsibility on to the last member of an unending series of particles can hardly be said to satisfy the mind which demands a clear physical conception of nature.

It may be objected, perhaps, that the difficulty just shown to exist in the case of the volumetric compression of an elastic substance does not obviously occur in regard to the compression of a rubber ball or of a spring, since in these cases it is merely change of shape, and not change of volume, which occurs, and it seems conceivable that resistance to change of shape might occur in structures built up of particles which themselves need possess no elasticity or compressibility. If, for instance, a chain of particles be grouped in a circle it seems natural that the circle should resist deformation—it being granted that particles in contact attract one another. Similarly there should be elastic resistance to the deformation of any symmetrical stable grouping. The force, however, of this argument is lost if the attraction is limited to the actual points of contact of one particle and its neighbours. A ring or polygon of bar magnets—each mounted, for instance, on a cork, and the group floated on water—does in fact resist deformation. The same would be true if the magnets were spherical and polarized on a diameter. In such a case, however, the attraction is not localized at the points of contact but is effected by the whole magnetic fields of the magnets acting for the most part between points of adjacent magnets not in contact. The phenomenon thus involves action at a distance. If, instead of magnets, a series of, say, glass balls were used, each ball smeared with mucilage so that adhesive force would be strictly localized at the points of contact, no force would be exhibited resisting deformation of the group.

The result, then, of our endeavour to show the simplicity of the conceptions of contact and of elastic action, has been to lead us to a position which is surely no less "absurd" than that of the man who accepts "action at a distance" without demur. If the conception of the latter is difficult that of contact and of elasticity is no less difficult.

Atomicity or continuity?

It may be asked whether these difficulties might not be avoided by adopting a theory of the continuity of matter instead of that of atomicity. If, however, matter were a continuum, and not built up of discrete particles, its conception would present the same difficulty which is so familiar in connection with the continuous series of numbers in mathematics. The series of numbers, 0, 1, 2, 3, 4, 5, proceeds in big jumps of 1. In between 1 and 2 we can write 1·5; in between 2 and 3, 2·5, and so on. The jumps are smaller but still big, and in order to make the series absolutely continuous an indefinitely great number of intermediate numerals would have to be inserted in each gap. The mind tries to conceive of such a series written out, but as any such conception presents a series of consecutive elements, it realizes that between each pair of these elements, *however numerous they may have become*, an infinite number of new elements must now be inserted. Thus the continuous series of numbers is impossible of conception.

Similarly the conception of a continuous medium presents the difficulty that between any particle in it and what we try to think of as the adjacent particle, there must lie an infinite number of other particles, or, in other words, no particle can be said to be adjacent to another particle.

Conceptions of this sort are surely as difficult as that of action at a distance.

What is a conception?

The consideration just outlined of the meaning of "possibility of conception" seems to lead to the conclusion that no scientific conception is possible! How can this be reconciled with the common acceptance of such terms as "familiar conception" and the classification of ideas into those that are conceivable and those that are inconceivable?

The answer appears as soon as the attempt is made to define "conception."

The mind of a human observer receives impressions through the various senses. These impressions succeed one another in time-order; they are at first a meaningless series of sensations but the memory automatically records them. As the series progresses repetitions occur; first one sensation and then others are repeated, then a particular series of repetitions is repeated and later a combination of series of repetitions. The memory records these repetitions and if a series of sensations has been many times repeated the mind "recognizes" this series, so that as soon as the first steps in it are again encountered the mind "expects" the completion of the series.

When, through long experience of the repetition of series of sensations many of which are of great complexity, the mind has become trained in the art of expectation the things expected become correspondingly complex. Such expectations are called "conceptions."

A conception is thus the mental expectation of what events are likely—in accordance with the lessons of previous experience—to follow certain given events.

When the given events which introduce an expectation or conception suggest a series of sensations such as many men have experienced or have pictured, the conception is called familiar: but if the given events are a series hitherto unknown they call forth no definite expectation and they warrant no conception.

The suggestion is, therefore, that when, in common parlance a conception is called possible or even familiar, the meaning is not that this conception has a basis capable of being understood, but merely that it is based on a series of events whose repetition in nature has made it familiar to most men.

The usefulness of "impossible" conceptions

An "impossible conception" is a suggested or tentative expectation based upon a proposed given series of events, this series being one, however, of the actual experience of which there is no record and whose "possibility" is therefore in doubt.

If the proposed given series of events bore no apparent relation to any series actually found in nature and whose practical consequences were known, the tentative expectation could be no more than a random guess and would have no value. It is possible, however, by proposing a series of events which appear to bear analogy to some known series, and basing on this an expectation formed in analogy with the realized expectation resulting from the known series, to arrive at a tentative expectation which may serve to suggest useful scientific research. This research may lead to the discovery of new facts justifying the tentative expectation, and changing it from an "impossible" conception into a possible one, which, in due time, may become familiar.

An interesting case in point is that of the 4-dimensional space-time continuum proposed by the Theory of Relativity. We have ample experience of series of events in the continuum known as 3-dimensional space and, for instance, the geometrical properties of such space can readily be described in mathematical symbols. It is easy to write down by formal analogy the geometrical properties of a 4-dimensional continuum. Thus, mathematically, the properties of a 4-dimensional continuum present no difficulty. Nevertheless, according to our definition these properties are impossible of conception because they are expressed in an argument based upon the assumption of the existence of a 4-dimensional continuum of space and time and this, though possible of expression in mathematical terms, is not regarded—except by Relativists—as definable in terms of actual experience.

Minkowski's "world"

Minkowski described this impossible conception and showed that the M.-M. phenomenon would form a suitable basis for such a conception. All recorded experience had been expressed in terms of 3-dimensional space as co-existent with but independent of 1-dimensional time. The conception that space and time, far from being independent fields for the ordering of events, were in truth the arbitrarily divided elements of a 4-dimensional continuum was an "impossible" conception, no basis for such a conception having yet been recognized in experience.

It was, however, a fact that certain observations did fail to fit in with any of the conceptions hitherto regarded as possible. They fitted in perfectly with Minkowski's "impossible" conception. The question thus was raised as to whether phenomena in general could be made consistent with this new conception.

The General Theory is an "impossible" conception

Einstein believed that they could and, on this assumption, built up his General Theory of Relativity. No such remarkable case of the justification of an "impossible" conception ever has been known before. All ordinary phenomena have proved, or give good promise of proving, consistent with it: phenomena—in particular the rotation of Mercury's elliptic orbit—that had been inexplicable have now become explicable and, most striking of all, this development by Einstein of Minkowski's impossible conception has led to the definite quantitative prediction of an unexpected—or at any rate only vaguely suspected—natural phenomenon, namely, the double-Newtonian bending of light rays in a gravitational field.

The probability now is, therefore, that the conception of the "world" of experience as a 4-dimensional space-time continuum is a "possible" conception. Should it prove indeed to be so, then it will surely become, in due time, a familiar conception.

The mixing of space and time

According to Minkowski's conception the "interval" between two events is represented by the space-time separation between two points in the 4-dimensional continuum. According to ordinary conceptions events are separated in two distinct ways, spatially and temporally. For instance the stroke of one o'clock on Big Ben and the stroke of one o'clock at Greenwich are said, ordinarily, to be synchronous in time and some few miles apart in space. According to Minkowski an observer, say, on the sun would record the matter differently (after having made accurate allowance for the time of transmission of the light—for we imagine the "strokes" to have been flashes of light) but with equal "truth." He would find that the two strokes were not synchronous and that they occurred at points in space a great distance apart.

The reality separating the events must surely be some definite quantity. It is not space nor is it time; Minkowski's conception makes it a mixture of the two—an "interval" in the 4-dimensional world.

Variously moving observers find various values for the space interval and various values for the time interval, but the Minkowski space-time interval, being the "real" measure of the separation of the events, is found by all observers to have one and the same measured value.

The difficulty of the idea of a 4-dimensional continuum

The simplest form of continuum may be represented by a line. A line is usually thought of as consisting of a series of points each of which represents one section of the line. A line is called a continuum because there are no gaps in it, that is to say it is possible for a point to move from any section of the line to any other section without going outside the line. The line is called a 1-dimensional continuum and it is said to be "generated by" the motion of a point.

Less simple than the line is the surface. A surface may be thought to consist of lines laid close together side by side, each line representing one section of the surface. A surface is called a continuum because there are no gaps in it, that is to say it is possible for a line to move from any section of the surface to any other section without going outside the surface. The surface is called a 2-dimensional continuum and it is said to be "generated by" the motion of a line.

Less simple than the surface is the solid volume. A solid may be thought to consist of surfaces laid close together, each surface representing one section of the solid. A solid is called a continuum because there are no gaps in it, that is to say it is possible for a surface to move from any section of the solid to any other section without going outside the solid. The solid is called a 3-dimensional continuum and it is said to be "generated by" the motion of a surface.

Thus a continuum consists of a continuous succession of components of the next lower number of dimensions.

By analogy we try to form a conception of a 4-dimensional continuum. It will have to consist of a continuous succession of 3-dimensional elements. Is this possible? We think of some solid body, say, a brick. In what sense is a continuous series of bricks a 4-dimensional continuum? A mere *row* of bricks is a continuous series but it is still no more than 3-dimensional. In building a row of bricks no new dimension is introduced; the row progresses in the same direction in space which already has been used to accommodate the thickness of each component brick, that is to say in one of the three spatial directions which belong to ordinary 3-dimensional space.

The continuous series in which bricks have to occur if they are to form a 4-dimensional continuum must be a series which progresses along a new dimension. A fourth spatial dimension is not available—it does not occur in the experience of any individual observer—and a brick having four *spatial* dimensions is nonsense. If, however, time may be regarded as a dimension it is possible that a brick may make use of time as a fourth dimension. In such a sense, then, a brick is a thing measuring 9 inches by 4½ inches by 3 inches in the three mutually perpendicular spatial directions defined by the three ridges which meet at one of its corners, and of indefinite length (namely, the length of the life of the brick) in the time direction. To make the three spatial dimensions independent they are measured along lines spoken of as mutually perpendicular, so that progress along any one of them does not constitute progress along any other of them.

To make the time direction independent of the three spatial directions, time must be measured along a line thought to be perpendicular to each and all of the three spatial directions. We cannot visualize this because to visualize means to form a mental picture, and as we already require time for all the mental pictures we make, whether the things pictured be of no dimensions, one dimension, two or three dimensions, time is not available for insertion as a fourth dimension into a mental picture. It is as though time, being a thing of one dimension is unable to perform two functions at once.

Mathematically the visualizing is easy enough—or perhaps is thought unnecessary—it consists merely in writing x, y, z and t (or x_1, x_2, x_3, x_4) instead of x, y and z (or x_1, x_2 and x_3).

We return for a moment to the analogous cases of fewer dimensions. The motion of a 2-dimensional continuum (a surface) parallel to itself produces a 3-dimensional continuum; that is to say, in order to generate the continuum of three dimensions a 2-dimensional continuum must be made to progress through the third dimension. By analogy, then, we may conclude that in order to generate a continuum of four dimensions a 3-dimensional thing must be made to pass through the fourth dimension. When—as in Minkowski's conception—the fourth dimension is time, the generation of the 4-dimensional thing corresponding to an ordinary 3-dimensional brick is brought about simply by making the brick pass through time—that is to say by letting it go through its own natural existence. And, after all, when we reflect that a brick has no chance to be a brick unless it be given time in which to have its being, the notion that, in truth, a brick is a 4-dimensional thing becomes almost obvious.

There are, however, defects in this line of reasoning. To speak of the generation of a line by the motion of a point is to use the word generate in a special conventional sense. One point cannot, by its motion or in any other way, generate a line in the sense of making it—an infinite number of points are needed for this. One plane cannot generate a solid—the solid needs for its existence the simultaneous existence of an infinite number of planes. By analogy, therefore, one brick cannot generate a 4-dimensional thing—an infinite number of bricks are required for this, all existing simultaneously and yet occupying consecutive positions along the fourth dimension, which we are trying to suppose is time!

A brick needs time in order to be a brick

The use just made of the word "simultaneous," meaning at one and the same time, indicates that time—to us as individuals —is not a dimension in the sense that length or distance is. A brick demands time in which to be a brick. Is it not equally true that a surface demands time in which to be a surface; and similarly for a line and lastly for a point? The no-dimensional thing called a point demands—in our individual conception of it—time in which to be a point. The mere existence of a point in time or, as we may express it, the passage of a point through time does not create a line: to pass through time is not to add another dimension.

The difficulty is only made greater if we try to maintain that even now, at this given instant, the brick exists simultaneously yesterday and to-morrow—for this is a contradiction in terms. The definition of the term "now" or "present" forbids it to include past or future.

The gradient of events

In some sense, however, our individual experience of the present does surely involve something of the past and something of the future. The mere instantaneous present has no meaning; the meaning of present experience seeming rather to lie in the process of change of which the present instant is a section. At every instant there is a gradient of events and it is this gradient, not the instant itself, which constitutes, or gives meaning to, present experience.

What then is the meaning of Minkowski's 4-dimensional continuum?

Can we define the difficulty more clearly? Whenever paradoxical or exasperating difficulties occur in the consideration of Relativity it is almost sure to be the universal observer's fault. In the present case he insists on our calling a brick a 4-dimensional thing. If he means a thing having four spatial dimensions his demand is absurd, but we must bear in mind that "spatial" dimensions are things not recognized by the universal observer; he no more would think of postulating four spatial dimensions for a brick than four temporal ones, for neither space measurements nor time measurements are realities to him. He postulates four dimensions; his world provides only one sort of dimension but there happen to be four distinct dimensions of this one sort: they can be measured in any desired "direction" provided the four dimensions are taken along mutually perpendicular directions—just as a brick retains its character in every respect no matter how its length, breadth and thickness be oriented provided they be kept mutually perpendicular.

In the world of an *individual* human observer every "object," whether it be a point, a line, a surface or a solid, needs time in which to "exist," and nothing of more than three homogeneous dimensions can exist, but in the world of the universal observer "existence" is without meaning, for, as we shall see later, the absolute nature of the order past, present, future—without which "existence" loses its meaning—is not recognized by the universal observer.

The universal observer's world is a static thing in which "time," being unoccupied in "passing," is free to act as a pure dimension, and measurement shows that the rôle it plays is indistinguishable from the rôle played by "space."

The "space-time interval"

Minkowski's proposal is not that time is a spatial dimension but rather that in the "real" world—the world of the universal observer—there are four dimensions all of the same nature. To an individual observer these four dimensions appear, three of them as spatial and the fourth as "time," but differently moving observers have different ideas as to which direction in the 4-dimensional continuum shall be called "time."

Experiments had led to the conclusion that different observers observing the same events and all possessing similarly constructed measuring instruments, but moving with different velocities relatively to the events, would find different values for the length intervals and different values for the time intervals necessary for the description of the events.

Whenever the space intervals found by two observers differed, the time intervals would differ also; in fact differently moving observers would act as though they had measuring rods of different lengths and clocks going at different rates.

Minkowski made the important discovery that by combining in a definite mathematical way the space and time records of any observer the result obtained was the same for every observer. The quantity resulting from this combination he called "space-time interval" or, briefly, "interval." The "interval" between two simple events ("point-events" as they are usually called) expresses their "real" relation, while the time interval and the distance between them are personal to each individual observer and do not depend "really" upon the events.

"Imaginary time"

The "interval" is expressed mathematically by

$$\sqrt{dx^2 + dy^2 + dz^2 + d\tau^2},$$

an expression obviously of similar form to

$$\sqrt{dx^2 + dy^2 + dz^2}.$$

The latter is the general expression for the distance between two points very close together in a 3-dimensional continuum in which three mutually perpendicular axes of reference x, y and z have been chosen. Thus the interval between two points in a 4-dimensional continuum is analogous to the distance between two points in 3-dimensional space. Minkowski found that if $d\tau$ in the above expression stood always for $\sqrt{-1}\,(dt)$, where dt is the small element of "time" separating the events, then the value found for the expression by different observers would be the same although, in a manner depending on their own motion relative to the event, these different observers would give to their systems of axes (x, y, z) different orientations.

In 3-dimensional space the distance $(\sqrt{dx^2 + dy^2 + dz^2})$ between two neighbouring points has the same value no matter in what direction the mutually perpendicular axes x, y and z be taken. The same is true, in 4-dimensional space, of "interval."

If the definition of "reality" demands that real relations between events shall be definable and therefore unambiguous—that is to say shall give the same results to all observers who choose to measure them—then length and time are not realities but interval is, and consistently with this we ought to say that the "real" world is not the ordinarily accepted compound consisting of a 3-dimensional space and an independent uniformly flowing 1-dimensional time, but is the 4-dimensional continuum in which the four—mutually perpendicular—axes x, y, z and τ can be taken. The fourth axis τ is the time axis, but to signify that actual time measurements before being plotted along this axis are to be multiplied by the square root of minus one, it is better to use the conventional word "imaginary" and speak of τ as the axis of "imaginary time."

Homogeneity of continua

The claim of the 4-dimensional space-time continuum to represent the "real" world, is based on the fact that when events are supposed to be located in such a world their relations are expressible in terms of "intervals"—quantities similarly measured by all observers—and so are "real." In no other sort of world that yet has been proposed can relations be defined which prove to be real.

The acceptance of such a 4-dimensional "world" depends, of course, upon our acceptance of the pragmatic definition of reality. All the advantages of the Theory of Relativity, as well as all the difficulties which its conception presents, are due to the form of this definition, but if we accept it then the Minkowski 4-dimensional continuum which its inventor called "world" is a world which, in respect of a sufficiently definable "reality" underlying the observable phenomena of nature, is common to all observers. In such a world—and in no other that yet has been invented—all observers are able to agree as to what "really" is happening. It may be possible to invent alternative worlds no less satisfactory and in all probability growth of knowledge in the future will make necessary at least a development and modification of the Minkowski world, but for the present the Minkowski world is, for scientific purposes, the least unsatisfactory. It is the world of the universal observer and most certainly is not the world in which I, as a man, have my being. But then my own world is useless for scientific purposes.

In the ordinary—pre-Minkowski—world, the length of a body and the time interval between two events are regarded as realities. This notion was acceptable in the old days when no case was known in which accurately made measurements of space and of time, even though made by differently moving observers, could give discordant results, but since the time of Fresnel (1818) and of Fizeau (1851) certain physical measurements have presented inconsistencies, and it was shown in 1895 by Lorentz and by Fitzgerald, that these inconsistencies might be expected if different observers, differing in respect of their motion relative to the events being observed, formed different judgments of space intervals and of time intervals. From this point on the reality of space and of time as independent things was held in doubt. Einstein boldly proposed to deprive space intervals and time intervals of the benefit of the doubt and to regard them, in accordance with Minkowski's theory—suitably developed to cover all phenomena—as things whose magnitude is imputed to events by the observer and not, strictly, as relations belonging to the events.

Based as it thus is, upon a definition of reality best suited not to any particular individual observer, but to the whole world of observers (or rather of conceivable observers, since the definition respects the composite view of observers whose positions and velocities are in many cases incompatible with sojourn on the earth), the Theory of Relativity demands a breadth of view outside the range of any individual. It is hardly surprising, therefore, that it presents difficulties to the individual mind, and that no man can visualize the "world" in which "events" consist of those "realities" with which the theory deals. All that any one individual man can visualize is the limited—and hopelessly complex—world of which he himself is the hub.

Thus Relativity expresses the view not of any one individual observer but of an imagined observer who has an eye and a finger in every corner of the universe. This universal observer is complex not only in regard to the unlimited number of his eyes and fingers but more especially of the unlimited variety of the relative motions these eyes and fingers can assume. This broad-visioned observer can assume and combine the points of view of observers located in all parts of the universe and moving, relatively to the observed events, with all possible velocities. The picture this observer gets is a composite one such as no individual can conceive, and it is supposed to represent "reality."

The superiority of the powers of such an observer in visualizing the world of experience may be compared with the superiority of our own powers over those of a man born blind in visualizing events in 3-dimensional space. Can such a man clearly visualize a brick? The mental vision he has of the brick is derived, probably, from the sense of touch, yet neither this sense nor that of sight is necessary to an appreciation of the events of experience—the phenomena of nature. Thus a direct visualizing of 3-dimensional space is not necessary to a sufficient appreciation of the individual human view of nature: neither is a direct visualizing of 4-dimensional space-time necessary to a sufficient appreciation of the universal observer's view.

4-dimensional space

The expression 4-dimensional *space* is a very old one, and the discussion of the supposed properties of such space is fascinating, nevertheless the only *space* of which we have experience is space occupied by things and things are 3-dimensional. To speak of 4-dimensional *space* is nonsense, just as it is nonsense to say that a brick has four spatial dimensions. The world in which a brick has four dimensions is the space-time world and that is not the world of individual human experience—it is the world of that "absurd" creature of the human mind, called the universal observer; a creature whom we have to tolerate because he is so useful.

The 4-dimensional continuum of space and time ought not to be called 4-dimensional *space*; to call it "space" is to misuse the word "space" so as to court confusion and misunderstanding.

Some confusion seems unavoidable because, in trying to explain the nature of space-time there seems to be no course available except to follow the analogy of continua of fewer dimensions, and the only familiar examples of such are those of 1, 2 and 3-dimensional space.

Space-time is, in one sense (namely, the sense understood by an individual human observer), a mixed continuum, one of whose dimensions (τ) is of different nature from the other three (x, y and z)—which themselves form the homogeneous continuum of the human observer's space.

Thus it might seem more appropriate to approach the problem of space-time by way of the analogy of some mixed continuum rather than by way of the homogeneous continua of space.

Examples of mixed continua

Consideration of a mixed continuum is instructive. We take, for example, the 2-dimensional continuum formed by the weights of men and their ages.

So far as we can tell a man's weight progresses not by sudden jumps but continuously and his age does certainly advance along the continuum called time. Hence if we record a man's weight at different ages by means of dots on a vertical sheet of paper, making the vertical height of a dot above a fixed horizontal line on the sheet proportional to the man's weight and the horizontal distance from a fixed vertical line proportional to his age, the points, if sufficiently numerous, should form a continuous curve. On such a sheet the age-weight curves of all possible men may be plotted, and the sheet represents the age-weight continuum.

The mixed nature of this continuum becomes evident when we try to express the meaning of the distance between two points, taken, for instance, on the curve of some particular man. An interval of one inch measured along the curve at that region which corresponds, say, to the man's birthday anniversary, indicates almost entirely increase of weight, while an equal interval on the day following indicates almost entirely passage of time. Thus the meaning of an interval of given magnitude depends upon the orientation of that interval in respect to the fixed axes of reference. The birthday interval was almost parallel to the axis of weight while the next day's section of the curve was almost parallel to the axis of age; hence, not only may equal intervals in different regions of the curve have different meanings, but the meaning of any given interval will be changed if, leaving the curve alone, the orientation of the mutually perpendicular axes be changed.

These peculiarities characterize all mixed continua but an important point must be noticed, namely, that there is no absolute criterion in regard to a continuum as to whether it is homogeneous or mixed. The nature of a continuum in regard to homogeneity is a relative matter: from the point of view of one observer it may be homogeneous, from that of another it may be mixed.

To an observer, if we can conceive of such, who regards change of weight as having significance identical with that of change of age, the age-weight continuum is homogeneous. A familiar example is provided by 3-dimensional space. From the point of view of a man (regarded as a *human individual*) on board a ship, space is homogeneous or is mixed according as this individual is a good or a bad sailor. In either case the path of the man is a curved line in 3-dimensional space and any interval of, say, ten yards on the curve represents the man's passage through ten yards in space. To the good sailor this means ten yards of a homogeneous continuum, but to the bad sailor it means a very complex thing, for this observer with his narrow human point of view considers that there is a fundamental difference between ten yards of choppy space and ten yards of smooth space: he finds space far from homogeneous. It is not merely that horizontal distances have to him a different meaning from oblique ones; the meaning of a space interval depends also upon the accelerations occurring while he traverses it.

Choppy space and smooth space

However, to both the good sailor and the bad, in their capacity not as human individuals but as scientific observers, 3-dimensional space is homogeneous.

The fact that the homogeneity of a continuum is relative has an important bearing upon the understanding of Relativity. We have argued that space-time is a mixed continuum. From the point of view of an individual human observer this argument is sound, but from that of the universal observer it is at fault.

The language of intervals

The meaning of an "interval" of given magnitude in the 4-dimensional world is independent of the position and of the orientation of the axes of reference. But the language of intervals is that of the universal observer, not that of any particular individual human observer. To the latter the space-time continuum is not homogeneous, for every such observer makes his own distinction between time and space, which to him is absolute. It is, to him, obvious nonsense, to suggest that time is in any way comparable with one of the three dimensions of what he calls space. According to the strict

Newtonian conception space and time do not form a continuum; they are two independent continua—that is to say space (the space of all time) is conceived of as, theoretically, completely explorable in no time, and time (the time of all space) as capable of being passed through without change of position, that is to say, in no space.

The scientist tries to put himself into the position of the universal observer and finds his reward in the discovery of a world more nearly capable of definition than any previous world (for every age has its own "imagery" and therefore its own world) has been. Relatively to a particular observer the events of which the world is composed have a meaning such as they have to no other observer—because each has his own peculiar judgment about space and time intervals. The reason why, in spite of this, science has been able to progress, is this, that in regard to the great majority of phenomena which, until recently, have been known to scientists, the ambiguity of measurements has been, quantitatively, of negligible effect.

It is worth while investigating the "world" of the universal observer because in that world the future—or rather the counterpart of what, in our world, is the future—is more clearly visible than in our own, while every event in that world has its recognizable counterpart in ours.

Prediction of the future

Prediction of the future can be never more than a probable expectation; we cannot know that what has happened many times before will be repeated, but we can regard it as probable. Expectation of future events has, at present, no other basis than past experience, and past experience is usable as such a basis only when it has been systematically and intelligently recorded. The record must be, above all, consistent and unambiguous, and the only system yet discovered which makes such a record possible is that in which the quantities recorded are space-time "intervals." The Theory of Relativity makes use of this system.

CHAPTER III

THE complete Theory of Relativity is called the General Theory; it includes the older (Restricted or Special) theory as a special case.

The demand for Relativity

The novelty of the General Theory and the suddenness of its coming—for it was developed single-handed by Einstein between 1905 and 1915—suggest that the need for such a theory had not long been felt. The position was, however, that long discouraged by the apparent futility of attempts to develop such a theory, scientists had made shift to do without it. Probably no man more thoroughly appreciated the need than did Newton, for certainly none came so near to satisfying it.

Analysis of phenomena suggested, long ago, that the world of common experience consisted of "things in motion"; further analysis led to the resolution of motion into two factors, space and time, and of things into the mysterious factors called particles of matter. The motion of matter was definable as velocity (including rest, as zero velocity) and time rate of change of velocity (called acceleration). In order to change the velocity of matter a factor called "force" was said to be required, and as the common notion of force appeared to be compatible with the statement that the force required to accelerate matter was directly proportional to the quantity of the matter (a quantity called "mass") and to the magnitude of the acceleration, Newton was able, without disturbing common notions, to put mechanics upon a definite quantitative basis by *defining* "Force" as equal to the product of Mass and Acceleration ($F = M \times A$). Mass was the name given to the "quantity of matter" in a body (whatever that might mean) and the arbitrary assumption was made that the mass of a "given body" was absolutely invariable. On these bases Newton developed his laws of motion—laws which, whilst almost ideally simple in form, are, for all ordinary cases, so remarkably accurate that, by their aid, mechanical phenomena of the most varied and complex nature may be predicted with a degree of certainty and precision which seems to raise mechanics above all other experimental sciences and make of it a branch of mathematics.

The laws of Newtonian mechanics

Newton was faced, however, with two outstanding difficulties, namely, those of gravitational and of centrifugal force. Neither of these forces could be called, strictly, a mass-acceleration, although both produced and were measurable in terms of a mass-acceleration.

Two masses permanently at rest relatively to each other—and possessing, therefore, no relative acceleration—began immediately to suffer acceleration towards each other when made free to move. The force postulated as producing this acceleration was called "gravitational" force. Newton discovered that the magnitude of the force was proportional directly to the product of the masses of the bodies concerned and inversely to the square of the distance between their centres of gravity. This law applied with apparent precision to all bodies—whether they were planets or mere dust particles—independently of the nature of their material as well as of that of any medium separating the bodies. In fact gravitation acted equally well through the vacuous spaces separating the sun from its planets, and through solid lumps of lead. There was, moreover, no evidence that gravitation required time for its transmission.

Since gravitation, like other forces, was measured by the mass-acceleration it would produce, it also was called a mass-acceleration. Thus the force known as a body's weight is written $(M \times G)$, where G stands for the "acceleration due to gravity," and the body's weight is supposed still to be $(M \times G)$ even when the body lies permanently at rest on the earth and possesses, therefore, no acceleration.

A book lying on a table has no velocity relatively to the earth, nor is it gaining or losing velocity; it has, therefore, no acceleration. However, since gravitational force is of no importance except when active, and the action it produces is always a mass-acceleration, it is practically satisfactory to regard gravitational force as a mass-acceleration. The absurdity of the point of view is obscured by its convenience.

Centrifugal force

A similar difficulty is presented by centrifugal force. I whirl a loaded sling and observe that a force is set up in the sling which stretches it. The direction of the force is clearly along the sling because the sling does not bend. I express the force as a mass-acceleration. Every text-book of mechanics shows how the acceleration may be calculated, proving that its magnitude is (V^2/R) and that, therefore, the centrifugal force is (MV^2/R), but does any text-book point out the absurdity of calling the force a mass-acceleration? The force is directed along the sling, hence, since $F = M \times A$ and M is a non-directed quantity, the acceleration, A, must be along the sling. But the stone in the sling has no velocity along the sling, nor is it gaining any, that is to say it has no acceleration along the sling. Thus the Newtonian explanation of centrifugal force involves us in the absurdity of having to say that the stone both has and at the same time has not acceleration along the sling.

The absurdity may be obscured by saying that the acceleration is not along the sling but rather is along the *direction* of the sling. The sling is regarded as remaining unchanged; its direction as suffering continuous change.

But how do we know that the "direction" of the rotating sling is changing? May not space itself, for aught we know, be rotating along with the sling? Or is such a question meaningless?

I have evidence only that the sling rotates relatively to my hand, and therefore also to surrounding objects. Must I then conclude that the centrifugal force is due to the sling's motion relatively to surrounding objects, and that the force would disappear if all surrounding objects were annihilated?

Newton's rotating bucket

Newton gave much thought to this question but is said to have satisfied himself too readily by a simple and by no means conclusive experiment. He suspended a pail of water by a string and spun it by twisting the string. After a while the pail and the water in it were found to be rotating together, the effect of centrifugal force being seen in the piling up of the water around the sides of the pail. The question whether this force could be due to the rotation of surrounding objects relatively to the water appeared to be answered in the negative by the fact that when the spinning first began—the pail then spinning practically independently of the water, since time was required for the friction of the walls of the pail to accelerate the mass of water—no piling of the water occurred. The bucket was thus rotating relatively to the water—just as, a little while later, all surrounding objects would be doing—yet there was no evidence of centrifugal force acting upon the water.

Newton is said to have concluded from this inadequate experiment, that centrifugal force is due not to rotation relative to surrounding objects but to "absolute" rotation—that is to say rotation relative to absolute space. At any rate Newton did conclude that rotations, and accelerations in general, were absolute.

Thus, absolute space, even when "empty," seemed to have at least one property, namely, that of creating forces. As soon as absolute space related itself to matter after the manner of a rotation—or, in general, an acceleration—forces began to appear.

This, however, was not the only property which, in accordance with Newtonian mechanics, had to be attributed to space. Somehow or other space had to convey energy from the sun to the earth, and much of this space was supposed to be devoid of matter.

Newtonian mechanics recognizes two forms of energy, namely, kinetic and potential. The former is possessed only by matter in motion, the latter may be possessed by matter at rest, the amount of the energy depending in this case solely upon the position of the matter considered, relatively to other matter.

The sun pours energy (radiation) upon the earth independently of any change in the relative position of sun and earth, so that the energy thus gained is not potential energy—as usually defined—and, as we know, it shows itself immediately in the kinetic form of heat.

Must we then attribute to space the property of conveying kinetic energy? If so we have to suppose that, during the time of the transmission—which, in the case of radiation coming from distant stars, is some thousands of years—the energy is located in "space."

This difficulty led soon to the invention of the Newtonian "ether," the supposed "medium" by whose undulations energy—in the form of radiation—is transmitted. It was not thought to be a "material" medium (though the meaning of the word had not then, nor indeed has ever been strictly defined) but all sorts of properties, such as rigidity and elasticity and density, analogous to those possessed by matter, had to be attributed to it.

Absolute rotation

The difficulties and inconsistencies involved in the ether idea have formed always a strong incentive to the development of some theory, such as that of Relativity, in which an identifiable ether would be unnecessary. Newtonian mechanics demands an ether—that is to say an "absolute space"—in respect to which accelerations, and especially rotations, may be defined. If I spin myself round like a top I *feel* centrifugal force, which, therefore, I say is "real." My surroundings give no evidence of having anything to do with the matter, and I refuse to believe that any motion whatever on the part of my neighbours, or even of all the suns and planets in the universe could cause my arms to fly out radially as they do when I "truly" spin. I conclude, therefore, that when I spin round I do something in respect to absolute space, not merely in respect to other bodies.

This, crudely put, seems to be the essential argument for the existence of an absolute space or ether. What has Relativity to say about it?

The Theory of Relativity takes the standpoint that such an ether is both unwarranted and unnecessary. Positions, times, distances, velocities, accelerations, masses, forces, energy; all these are relations of things to an observer; not one of them is "real" or absolute. A mere lonely particle, isolated in unlimited empty space, could not be said to possess position or velocity or energy.

The theory assumes that if I stood quite still and the universe began to rotate about me, I should feel all the sensations of centrifugal force. Newton's experiment of the spinning bucket was obviously inadequate because the bucket itself formed an inappreciable speck of matter in comparison with the universe. The piling up of the water occurred when the whole universe (except the negligible bucket) rotated about the water and it was this relative motion which caused the centrifugal force.

It probably is impossible to contrive any practicable experiment by which the question of the relative nature of rotations can be put to conclusive test. It has been suggested that a rapidly rotating massive flywheel should create a field of centrifugal force great enough to produce some detectable effect upon a delicately suspended balance placed near the wheel, but the probability of success is too small—since even the largest practicable flywheel is of infinitesimal mass in comparison with the mass of the earth, to say nothing of the universe—to make the experiment worth while.

Although, therefore, it may be repugnant to most minds to suppose that the sun and stars are directly concerned in the process by which I reset my clinical thermometer or flick a fly off my nose, there is no scrap of practical evidence against the supposition, and the Theory of Relativity accepts it—at any rate tentatively. According, then, to Relativity, the phenomena of nature can be described completely in terms of things and the relative motions of things, there being neither ground nor need for introducing an "ether" of "absolute space." Among the phenomena of nature are forces, the foremost of which are those called "inertial" (true mass-accelerations), including as a special case centrifugal forces (for, as force is a Newtonian conception we use the conventions of Newtonian mechanics and neglect the inconsistencies involved), and those called "gravitational."

The assumption made by Relativity that the phenomena connected with, say, a spinning top, are described equally as well from the point of view that the top is at rest while the universe moves around it, as from the more usual point of view, namely, that the top is "spinning," implies that centrifugal forces—and the same argument may be extended to all other so-called inertial forces—are due directly to the relative motion of bodies. The notion involved is unusual, but can easily be grasped. If, however, the Theory of Relativity is to be General, it must ascribe all forces, including gravitational ones, to the relative motion of things.

A Relativist would hardly state the matter in this way, for, as will be shown later, Relativity recognizes neither "force" nor "motion," as such; but it is practically impossible for people brought up in Newtonian principles to enter the world of Relativity with any confidence unless they have first surveyed it from without through Newtonian glasses. In taking such a survey the phenomenon of gravitation presents formidable difficulties.

Gravitation is regarded, in Newtonian mechanics, as a static phenomenon. Two bodies, relatively at rest, exert mutual attraction—and we are accustomed to believing that they would do so even though they two were the only bodies in the whole of unlimited space. It seems, indeed, a contradiction in terms to identify gravitational force with relative motion.

Static forces

Does not, however, the very idea of a "static force"—in spite of the familiarity and convenience of the expression—involve contradiction? If force is real it must be something we can measure, and there is no known means of measuring a force except by letting the force become active. Gravitation is called a force because, when allowed to do so, it accelerates bodies. For convenience we speak of the force as "existing" even while it is prevented from action, but is it not, strictly, nonsense to speak thus? The Newtonian definition: "Force is that which changes or tends to change a body's state of rest or of uniform motion in a straight line," buries this difficulty under the convenient word "tends." A thing moves or it does not move. How can we define a mere tendency? One thinks of a wound-up spring—it surely "tends" to unwind. But how do we know this? Only because past experience reminds us that wound springs, when freed, usually do unwind. This makes of a "static force" a mere expectation—not a force at all. It is a force only when it is a true mass-acceleration. The force of gravity on a body has indubitable existence only when the body is actually falling. Yet how can this be reconciled with the fact that it is precisely while a body is actually falling that the body "experiences" no force?

An illustration commonly used in this connection is that of a freely-falling lift. A man in such a lift would exert no pressure on the lift floor; his hat would not press upon his head—in fact, if he were to raise his hat, hold it thus for a while and then release it, the hat would remain poised in space. In other words, no object in the lift would experience the force of gravitation.

In face of this difficulty one is inclined to go back, and say that the gravitational force truly exists (*acts on* a body) only when the body is *not* falling—for only then does the body "experience" the force.

To relieve the dilemma we ask, what does the expression "to experience a force" mean? A body (say, a book) lying at rest on a table "experiences" the force of gravitation. Its lower leaves are more compressed than the upper ones, and, on account of this are, no doubt, a little thinner than the upper leaves. But change of thickness is not force. Where is the "reality" of the force which the book "experiences"? We say the phenomenon of compression exhibits the force. But how do we picture the inwardness of this phenomenon? The molecules of the paper are in relative motion—the average velocity of the average molecule being determined by the absolute temperature of the paper. Gravitation does not alter the temperature, nor does it affect, therefore, the average velocity of the molecules. Gravitation presses the molecules more closely together, that is to say it reduces the spatial limits within which the molecules vibrate: all that it does, then, is to change the paths, that is to say modify the motion, of the molecules. The point to notice is this, that the modification which gravitation produces in the motion of the molecules must be precisely that which *will prevent the molecules from "experiencing" gravitational force.*

Our dilemma seems, in fact, to become only worse as the argument proceeds. To say that gravitational force is exhibited in the phenomenon of a body allowed to fall freely, seems to be nonsense as soon as we reflect that it is just such a body which experiences no gravitation, while to say, on the other hand, that gravitational force is exhibited in the case of a body lying at rest on the earth is either, from the point of view of the body as a whole, to make of the force a mere expectation, or, from the point of view of the body as a molecular system, to make of the force a "suicide," a thing whose only effect is to destroy itself.

What does all this confusing quibble amount to? It simply removes the camouflage which, since the time of Newton, has obscured the absurdity of the idea that "force" is a reality.

Relativity discards "force"

Relativity discards the idea of the existence of forces. All that it recognizes is this: Particles of matter have "natural paths" about the nature of which nothing but experiment can tell us. The experience of ordinary human observers, watching and describing ordinary events, has led to a plausible convention, namely, that the natural path of any particle is a straight line. Observers have been sometimes ignorant, sometimes negligent of the fact that each observer has his own peculiar definition of "straight"; and when any particular observer has found a particle whose path is not along what he himself calls a straight line, he invents a "force" to account for the irregularity. He forgets that, in every such case, it is possible to specify another observer—to whom suitable motion has been attributed—who will find the path of that very same particle to be rectilinear, and who will say, therefore (and with equal right), that the particle is moving freely, unaffected by any force. Had he remembered this he would surely have realized that force—regarded as belonging to the system observed—is a mere fiction. The force he has attributed to that system, and which he has foolishly regarded as a factor necessary to the description of that system, is merely an expression of the relation of the observer to the system.

These considerations put a new aspect upon the problem which Relativity has to face in ascribing forces—including gravitation—to the relative motion of things. Forces are factors in the relation of particular individual observers to the systems they are observing; differently moving observers having, usually, different force-factors in their relations to one and the same system. Relativity is not concerned with the particular points of view of individual observers, its concern is reality, whose factors are those experiences, and those only, which occur in the (corrected) observations of one and all observers, however diverse may be their movements in relation to the events being observed—and reality is described in the record of the universal observer.

The universal observer regards forces as fictitious because he finds that they are incapable of satisfactory definition—their definition, and even their existence, depending solely on the point of view assumed by the definer, and there being no criterion by which one point of view may be regarded as more "true" than another—and he banishes the word "force" from his description of reality. He contends that if he succeeds in expressing that which there is of "reality" in the space-time relations of things there will be nothing in nature that he can have overlooked.

Before following the universal observer further, it will be well to deal with some of the difficulties which, in spite of what has just been said, can hardly fail to cause hesitation.

Spinning top or rotating universe?

1st. Can nothing more be said to reconcile the mind to the theory that if the universe rotated around a top, all the phenomena, including the ultimate bursting of the top, would occur which we know do occur when the top spins?

An effective demonstration of the reasonableness of the contention consists in showing how very unreasonable is the position of one who disputes the contention.

Let it be proposed that if the universe rotated around the top, the latter would experience no centrifugal effect—neither bulging around its middle nor ultimately flying apart.

The mind must then be willing to make two pictures, *A* and *B*, of the universe—which we define as embodying the whole of nature. Both pictures contain the same elements, namely, the top, the earth, the sun, the stars and all other things.

In *A* (the top "spinning") these things—including the molecules of the bulging top—are moving relatively to one another in the manner natural to them.

In *B* (the top "at rest," while the other elements of the universe rotate around it) we have the same things as in *A*, and the same relative motions—*except that the molecules of the top are moving in slightly different paths* relatively to surrounding bodies, since, by the proposed assumption, the material of the top is undistorted by centrifugal force.

Can the mind reconcile itself to the possibility of two such pictures? If so, how does it account for the difference between them? The phenomena in *B* differ from those in *A*, and minds usually demand for different phenomena different "causes." What can the mind suggest as the cause of the difference between the two pictures? The search is vain unless one postulates some cause outside the picture, but this means outside the universe and contradicts our assumption that the universe embodies the whole of nature. A similar difficulty is raised by any alternative construction of the picture *B*, other than that which makes it identical with *A*. Thus the mind finds itself in a dilemma: the Relativity theory as applied to rotation is repugnant, and any alternative theory involves contradiction. The mind will surely do well to throw off its repugnance.

The problem of the two apples

2nd. The ordinary text-book explanation of centrifugal force may, in the strict sense, involve an absurdity, but minds ordinarily accept it with satisfaction. If Relativity proposes to put centrifugal force and gravitational force into one and the same class, why cannot it give to the mind some treatment of gravitation which the mind may accept as readily as it does accept the ordinary treatment of centrifugal force?

Gravitation presents a much more formidable difficulty than does centrifugal force, for two reasons; first, because while centrifugal force acts solely in the plane of rotation of each particle of the rotating body and so may be attributed without unreason to the rotation and this, in its turn, to relative motion in respect to surrounding bodies, gravitation acts simultaneously in all the planes which contain the centre of gravity of the body—and, obviously, no sort of motion or relative motion the body may have can possibly possess components in all such planes at once. And as it is contrary to all accepted notions to suppose that a motion can have an effect—such as the production of a force—except in some direction in which the motion has a component, it seems absurd that gravitational force should be attributed to relative motion.

This difficulty may be expressed in another way: Along any diameter of the earth, as, for instance, that which joins its North pole (N) and its South pole (S), the forces of gravity act in opposite directions, for at N they act in the sense N to S, and at S in the sense S to N, so that, for instance, two falling apples, one near N and the other near S, travel towards each other, their mutual approach being an accelerated one, the mass-acceleration of each apple measuring the gravitational force involved.

Relativity says it is equally as legitimate to speak of the motion of the earth towards the apples as of the motion of the apples towards the earth. But there are two apples and only one earth, and while, obviously, the apples can possess, simultaneously, oppositely directed motions, the earth can do so only by expanding or by coming in halves—which experience shows it never does.

Can a body move two ways at once?

Neither Newton nor Einstein has pretended to *explain* gravitation; merely they have tried to describe what happens. The Newtonian description seems straightforward enough, but it postulates the mysterious thing called gravitation, in addition to space, time and things. According to this description it is true to say that both apples move towards the earth, and absurdly untrue to say that the earth moves towards both apples. The absurdity consists in supposing that the earth can move two ways at once, namely, in the direction *S–N* in order to meet the one apple and, simultaneously, in the opposite direction *N–S* in order to meet the other. The absurdity needs, for its demonstration, no other bodies besides the earth and the two apples. But does it not need "absolute space" as well? Would not the ground of the absurdity be removed if we were to discard the tacitly, and perhaps unwittingly assumed existence of "absolute space"? Why cannot the earth move two ways at once? What does "two ways" mean? It surely means "two ways in space," and involves the assumption that "ways in space" are capable of being—if not identified individually—at any rate distinguished one from another. Can an example be given of two ways in space? We have just used as such the "two ways in space" defined by the two radii *SO* and *NO* of the earth (*O* being the earth's centre of gravity). But the earth is not "space," it is "matter."

Thus the fact begins to emerge that "motion in space" is merely a figure of speech, the attempt to define it leading always, and of necessity, to a statement of relative motion of matter.

We face again the repugnant notion that the earth does indeed move towards both the apples at once. Unless biassed by convention neither apple, speaking for itself, can raise any objection to the contention that itself is at rest and the earth is coming to strike it. Such a point of view is, indeed, a very reasonable one, for the apple's bodily sensations tell it there is no force acting upon any part of it. The notion that its own body is being accelerated towards the earth, is repugnant to the apple, for, by Newton's iaws—the truth of which every educated apple lives chiefly to demonstrate—acceleration would involve force, and the apple experiences no force.

Thus the apple *A* is happy to accept the point of view that the earth is rushing up to strike it. The same would be true, at the same time, also of *B*. But what of an external observer contemplating the three actors, *A*, *B* and earth (*E*), in this event? What can he say to the statement that *E* is moving simultaneously towards both *A* and *B*?

The description might at first seem repugnant on the ground that the earth would have either to split or to expand, but as soon as he realized that the earth's motion simultaneously towards *A* and towards *B* was accompanied by a decrease in the distance *AB* exactly sufficing to make splitting or expanding unnecessary, he would see that his mental repugnance had been unfounded, and would realize that the world could be regarded, legitimately, as moving simultaneously towards any desired number of falling apples, situated in random positions around the earth.

This point of view is so unconventional and yet so crucial to the happy acceptance—even tentatively—of the Theory of Relativity, that it may be well to discuss still another example. In principle it will be the same as that of the apples, but the example will bring out more definitely the fact that it is our tacit, though unwarranted faith in the existence of a unique "absolute space" or "ether," in regard to which rest or motion may be defined, which creates the difficulty.

A man sitting on a rock finds that he can get up and walk to a distant tree. Half in fun he puts to himself this question: Am I sure that it is I who move towards the tree and not, rather, the tree that moves towards me, all that I do, meanwhile, being to make the leg motion necessary to allow the earth, bringing with it the tree, to slide under me? The idea is repugnant because I find it incongruous that my little effort should actuate the tree, and the earth, and the universe, calling into being the immeasurably great force required thus to impart motion to them.

Yet while I pause to admire the clinching nature of this point about the immeasurably great force, and to think how admirably Newton's laws are coming to my aid in the search after truth, an awkward rejoinder—based also on Newton's laws—occurs to me. The problem concerns two factors, namely, myself and the remainder of the universe. Do I move towards that spot in the universe I call the tree, or does the universe bring its tree to me. The force required to impart to the immeasurably massive universe the motion necessary to bring the tree to me in the time occupied in the journey, is a force too absurdly great to be called into being by my own effort; the fact must be, therefore, that it is I who go to the tree, for the force required to give to my body the necessary motion is of reasonable magnitude. But what about Newton's law that action and reaction are equal? I and the universe are the two things concerned, and no other things exist, hence when I and the universe approach each other, the force ("action") bringing me towards the universe must equal the force ("reaction") bringing the universe towards me,—that is to say a very small force of action must equal, by Newton's law, the immeasurably great force of "reaction." Thus my clinching argument involves an absurdity.

It may be objected that Newtonian mechanics must not be blamed for the absurdity just proposed, an absurdity which is avoided by postulating that both I and the universe move; not one of us alone. As my body's mass is small and that of the remainder of the universe is very great the force of *action* moves me through a great distance, while the equal force of *reaction* moves the universe through a small distance. But the force of this objection depends upon the assumption that I and the universe may have motion which is *absolute*, and not merely motion relative to each other. It is this assumption which Relativity holds to be unwarranted.

In disgust I return to my rock, and over there, away beyond the tree, I see another man sitting on another rock. As I rise and walk to the tree it happens that he also begins to do so, coming towards me as we both, from opposite sides, approach the tree. If it be held legitimate for me to maintain that the tree is moving towards me while I remain "fixed," it must be held equally legitimate for the other man to maintain that the tree is moving at the same time towards him, and that he is "fixed." However, as, in fact, we two men are all the while approaching each other, it is absurd to say we both are "fixed." Here at last I surely have a proof that it is I who come to the tree and not the tree that comes to me.

Let us examine the proof: it is a *reductio ad absurdum*. Are we compelled to admit the absurdity? On what ground do we call it absurd that two men, though in relative motion, should be at the same time both "fixed"? Any two men, in relative motion of whatever kind, are at the same time "fixed" —in relation, say, each to his own head. Or again, if two men move around a circular track, they must both remain "fixed" relatively to the centre of the track, no matter what their motion relative to each other may be. Thus it is not necessarily absurd to say that two men may be in relative motion yet at the same time "fixed." All depends upon the meaning of "fixed," for this is a relative term. Relative to what are the men fixed?

It would be absurd that while fixed relative to, say, the floor of a room two men should be said to be at the same time in relative motion, and, in general, the absurdity must be admitted provided, and only provided, the term "fixed" is used relatively to one and the same rigid, identifiable frame of reference.

Thus we can admit the "proof" in question only provided the term "fixed," as applied to the two men, is used in respect of one and the same, identifiable, rigid frame of reference.

What was meant by saying the men were "fixed," while the tree came towards them? In reference to what were the men "fixed"? The conventional answer is; in reference to "space" or the "ether of space." Accepting this answer, the *reductio ad absurdum* must be conceded provided, and only provided, the second man was to be regarded as fixed in reference to the "ether," and the latter was truly a rigid identifiable frame of reference.

Thus the argument depends upon the assumption that there is such an "ether"—that is to say an "absolute space."

The Theory of Relativity has been proposed because, in spite of long years of endeavour, scientists have been unable to substantiate the existence of a rigid identifiable ether.

Each man has a right to his own special ether

If, as seems likely, men will continue to use words which imply the existence of absolute space, they must, in fairness to their brothers, be broad-minded enough to concede to every man an equal right to dogmatize as to what is "fixed"—in other words each man must be allowed to possess his own personal ether, which pervades the universe. To concede this is to concede the existence of an unlimited number of rigid ethers, all co-existing, and having all conceivable sorts of motion relative one to another. Such a notion, fantastic though it may seem, is not more inconsistent with experience than is the vague notion of the unique ether, fixed to the stars.

CHAPTER IV

The General Theory of Relativity—its point of view

WE proceed now under the assumption that the point of view of the General Theory of Relativity has been accepted. To put this viewpoint more explicitly we may state it thus: The whole body of natural phenomena is capable of complete description in terms of things and the relative motion of things. In order that the description may be made in terms which have long been in ordinary use, it may express motion in respect of some system called "fixed," but, according to the theory, the choice (provided it be consistently adhered to throughout the description) of what system shall be called "fixed" is arbitrary. One description may assume as fixed the sun, another the earth, a third the apple A, a fourth the apple B, or, out of mere perversity I may assume that the one and only thing in the universe, which shall be regarded as fixed is the chair in which I am sitting. The form and wording of the description will depend upon which system has been chosen as fixed, but all these descriptions are equivalent and all are complete.

The falling lift

One or two of the interesting consequences of accepting the Relativity point of view are fairly obvious. For instance: I am in a descending lift, and with a piece of chalk I draw a straight line on the wall of the lift parallel to the floor. The lift wall happens to be made of wire gauze so that my chalk, while making a "horizontal straight line" across the gauze, passes through the perforations in the gauze and marks also the wall of the well down which the lift is travelling. On this well-wall the mark will be:

(*a*) a straight line *if the lift's velocity is uniform*—the inclination of this line to the horizontal being a measure of the lift's velocity;

(*b*) a curve, if the lift's velocity is not uniform: for instance, if the lift were falling freely the curve would be (approximately) a parabola with vertical axis—like the path of a bullet fired horizontally with a horizontal velocity equal to that with which I drew the chalk from left to right across the lift wall.

We have, then, various alternative descriptions of the chalk's path:

(*a*) From the point of view that the lift is fixed the chalk's path is a horizontal straight line.

(*b*) From the point of view that the well-wall is fixed, relative to which the lift has uniform vertical velocity, the path is an oblique straight line.

(*c*) From the point of view that the well-wall is fixed, relative to which the lift is suffering uniform downward acceleration, the path is a parabola.

Relativity says that none of these descriptions is more legitimate or true than another and that they are equivalent, their differences being of no importance since they result solely from the introduction by the describer, of unnecessary hypotheses in regard to what bodies are "fixed," hypotheses which express the unwarranted assumption of the existence of a Newtonian ether.

In other words; though observers who describe the motion of a body will insist on specifying the "shape" of its path, shape is, in reality, a meaningless term—one observer calling it straight, and another calling it bent, and neither being better justified than the other.

And just as the bending of the chalk's path is an unnecessary fiction, so also is the "cause" of that bending, namely "force"—in this case the so-called force of gravitation. The Theory of Relativity aims at descriptions from which such terms as shape and force have been eliminated.

Shape and size

The broadening of view called for in the acceptance of the principle of Relativity goes far beyond what has just been indicated.

It has been proposed that any desired system may be used, legitimately, as the "fixed" system of reference on which to base a true description of our world of things in motion. Naturally, we suppose, the chosen fixed system must be a rigid one. But is such a supposition necessary? A rigid system is one whose shape is fixed, but we have agreed that—at any rate in describing paths—shape is a relative property, not a "reality." And it would seem anomalous to attribute reality to the shape of, say, a solid body, while the shape of any one of its profiles (since these are *paths* in space) is regarded as unreal.

In addition to fixed shape a rigid body must have fixed size, and if our "fixed" system of reference need have no particular shape, why need its size be specifiable? Does not, indeed, the discarding of shape imply the discarding of size also? The "fixed" system of reference might consist, for instance, of three particles of matter. The shape of this system is triangular. Would it not be anomalous for Relativity to demand that while we may alter as desired the shape of this triangle we must be careful not to change its area?

Or, again: to assign "reality" to size means to assign reality to distance in space, and surely if absolute space is meaningless so also is absolute distance in space. The difficulty about regarding both the apples *A* and *B* (p. 72) simultaneously as "fixed" arose from our knowledge that their distance apart was changing, and our tacit assumption about space, that if two things were fixed in it then the distance separating them must remain constant. Thus the assumption that fixity in space was absolute implied that distance in space was absolute also.

The chalk path made in the lift was a straight line, say, 3 feet long as seen by the lift-man, but as marked on the well-wall the path was a parabola of much more than 3 feet in length—say, 5 feet. Relativity regards the "3 ft. straight line" and the "5 ft. parabola" as equivalent descriptions of the chalk's path, there being no "reality" in their difference either of shape or of size*.

From this point the endeavour will be to show how Relativity does consistently accept the consequences of discarding the Newtonian ether (by which always is meant a rigid, identifiable ether) even though this makes fiction not only of "shape," but also of rigidity and size, and that, in making the claim that any arbitrary system of reference is legitimate as a basis for a true and complete description of nature, no restriction is made as to the size or the shape of the system, it being permissible, indeed, to regard the reference system as a "mollusc," a jelly-like thing whose size and shape may be, if desired, continually varying in an arbitrary fashion, even while the description actually is in progress.

---

* We shall see later that the path itself, as an isolated path of a mere particle, is without meaning, but this need not delay the present argument.

The mention of the "mollusc"* calls for some words of reassurance. It is one thing for a mere irresponsible poet, working with the stuff that dreams are made of, to call the world his oyster, but it is a different thing for a serious scientist, working with the stuff that facts are made of, to call the world his mollusc.

Fortunately reassurance can be found in a few considerations which are both simple and fundamental—considerations which hardly would need mentioning were it not that the very obviousness of the truth underlying them had so clashed with convention that it had been found expedient to forget if not actually to discredit them.

Does time flow uniformly?

What is meant by saying that time flows evenly? Simply that each minute is of the same length as every other minute—and this is because each second is like every other second, while this is because a pendulum makes *equal* beats. Experiment shows that the time of beat of a pendulum depends only on the pendulum's length and the intensity of the gravitational field in which it swings, and knowing that these factors are constant we are sure that the beats are equal. We know, by measuring it, that the pendulum has constant length. But suppose I suggest that every day, between certain hours, all the linear dimensions of all things in the universe, including pendulums and measuring rods, become one quarter of their usual values: can anyone disprove this? Obviously not by means of a measuring rod. My watch, however, should settle the matter, by showing that the shortened pendulum was making twice as many beats to the minute as it used to. But this assumes that my watch still measures true minutes. Actually the radius of its balance wheel will have been quartered and it seems reasonable to suppose that the watch, like the pendulum, may be going twice as fast as it used to go.

---

* See Einstein, *Relativity*, p. 99 (Methuen, 1920).

If it be replied that in this case the long hand of my watch will make, between two consecutive sun-risings, 48 circuits of the dial instead of 24, I point out that to say this is to neglect our assumption that the whole universe was to suffer contraction; and that since, as science tells us, the laws governing the pendulum are the same as those governing the planets, it seems certain that the speeding of the pendulum's beats will be accompanied by an equal speeding of the earth's rotation.

Other more subtle points may be raised, but the problem is an old one which often has been carefully discussed, and the conclusion has been reached that, in every detail of nature, compensation would be, almost certainly, complete. In other words, the uniformity of the passage of time is no more than a convenient fiction—we have no means of knowing anything about the manner of the passing of time or of the behaviour of the linear dimensions of so-called rigid bodies.

We have thus no ground for attributing "reality" to either time intervals or length intervals. It is merely by convention based on convenience in ordinary practical affairs that we assume the constancy of measuring rods and of pendulum beats.

The fact that upon these assumptions a sufficiently consistent and reliable system of practical life, with its elaborate schemes of science and industry, has been developed, goes far to warrant the "truth" of the underlying assumptions. If "truth" be "that which works" what, indeed, can be truer than the assumptions of the constancy of standards of length and time, since upon these assumptions all the activities of life have been made successfully to work?

The reply to this strong argument is to concede that *in so* *far as it works* the assumption is warranted; namely, it is warranted in all the ordinary affairs of life. It is on account, and solely on account of the extraordinary cases in which the assumption does not work, that Relativity proposes to investigate the consequences of doubting its legitimacy.

The outcome of the investigation is this, that although ordinary phenomena are not appreciably inconsistent with the assumption of the constancy and unambiguity of space and time standards, the assumption is one which, though plausible, is in no case necessary, while in the extraordinary cases referred to it is simply wrong.

Rejecting the assumption, Relativity has been able to find a new method of describing phenomena; a method which applies equally well to the extraordinary and to the ordinary phenomena. Relativity assumes that the world of space and time is indeed a mollusc-like structure, and this assumption is found to work better than the other, no phenomenon being known to be inconsistent with it.

The mathematics of molluscs

The problem of the mathematical description of world-line structure in a mollusc-like 4-dimensional continuum is very difficult, but for a reasonable understanding of the "way of looking at things" known as Relativity, only a general idea of the meaning of mollusc mathematics is necessary. Such an idea may be obtained by aid of analogy. We begin by outlining a simple mathematical method much used for describing events in the Newtonian world: it is known as the graphical method.

The flight of a bullet may be described by stating, in order, all the successive pairs of values of the bullet's vertical height above, and horizontal distance from, the mouth of the gun. The position of the gun's mouth having been indicated on a sheet of paper by a dot in the left-hand bottom corner of the sheet, a straight line through this point parallel to the bottom edge of the sheet will be used to represent zero-level and a series of equi-distant lines drawn parallel to this base line will represent successive equi-distant higher levels. Similarly, by a system of equi-distant vertical lines distance measured horizontally from the mouth of the gun will be indicated. In order to record the fact that the bullet passed through a point in space, say, 50 ft. above zero-level and 600 ft. horizontally away from the gun, a dot is made on the chart at the point of crossing of the horizontal line representing a level of 50 ft. above zero and the vertical line representing 600 ft.

Paper prepared for use in this way is called "squared paper," because the lines upon it divide its surface into uniform square meshes, but it is interesting to note that, for such a purpose as recording the progress of a bullet, the lines on the sheet need not be straight, nor yet mutually perpendicular, nor even uniformly spaced. For suppose a photograph were made of the chart and the photographic film were to get stretched or distorted in any way; the crossings of the lines would be unaffected and the numerical order of the lines could be read off as before, so that if a dot had been made upon the chart at the point where the horizontal line numbered 50 ft., crossed the vertical numbered 600 ft., this dot would still be seen at the same crossing. The lines might now be curved in arbitrary fashion and their spacings be arbitrarily irregular, nevertheless, provided it had no tears nor folds, the film would be just as true a record of the bullet's progress as was the original chart. Merely the "shape" of the curve would no longer appear to the eye as parabolic—though when read in the light of the numberings upon its scale it still would represent a parabolic path. In fact, provided the chart, together with its numbered scales, is distorted as a whole the difference between the distorted and the undistorted chart is merely formal.

To the Relativist, however, who realizes that shape is an unreal property, and that no absolute meaning attaches to equality of spacing—whether this be indicated by actual length intervals on the chart surface, *or by the numbering of the lines*—the distorted film is in every way as complete and true a record as the original chart, even though the distortion of the film affected only the mesh lines and not the curve of flight.

Conversely the curve of flight, remaining always upon, say, an undistorted sheet of squared paper, may be arbitrarily distorted and stretched: to the Relativist it is just as true and complete a record of the bullet's flight as it was before!

But this seems to reduce to a farce the taking of the record of the bullet's flight. If any random curve whatever is as "true" a representation of the bullet's flight as was the original parabola we surely must conclude that the flight is an indefinable and meaningless event.

Such a conclusion is almost warranted but not quite. The space flight of a bullet, regarded as an isolated particle, is meaningless, but then we have no experience of an isolated particle. It is true that the "length intervals" covered by the bullet are without definite meaning, but the space progress of the bullet has certain absolute characteristics. The bullet always goes onward in space and its progress is continuous. These conditions are the expression of the continuity of the space curve of flight and of the fact that it never bends back and that it contains no knots. With these restrictions the distortion permissible in the curve of flight is arbitrary.

A more complex case may be considered. Let us suppose the flight of a bat were being described. A bullet's path lies always in a vertical plane; it makes use, therefore, of only two dimensions in space, but a bat moves sideways as well as up, down, forward and back; its path could not be recorded completely on a sheet of squared paper, the corresponding record would require a 3-dimensional chart, as, for instance, a cubical block of glass divided by a system of mutually perpendicular embedded threads into uniform cubical cells. A series of dots placed at the proper thread crossings would then describe the bat's path.

According to Relativity the curve of flight might be represented with equal truth and completeness by any continuous series of dots within the undistorted glass block, or, conversely, leaving the curve of flight unchanged, any arbitrary distortion of the mesh system is permissible—the condition being always respected that the curve of flight shall be continuous. Indeed we may carry to the very limit this liberty to distort the curve of flight arbitrarily, and may reduce the curve to a mere point! When Relativity says it is legitimate to distort a curve the meaning simply is this, that one observer's point of view is as good as another's—for the shape of the bat's curve of flight depends upon the motion of the observer relative to the bat. To say the curve of flight is a mere point is to take the point of view of an observer moving with the bat—and this point of view is as legitimate as any other.

We have now to consider the introduction of the time element into the curves of flight. Each point on the 2-dimensional space-curve of the bullet's flight ought to have a number attached to it indicating the time interval that had elapsed since the bullet left the gun's mouth, or, better still, the curve ought to be embedded in a block of glass, the distance of each point on the curve, measured from the front face into the thickness of the glass, indicating the time interval.

In this case also, since time intervals as well as space intervals are "unreal," distortion of the curve within the block, or, alternatively, distortion of the mesh-system while the curve remains unchanged, is without effect on the completeness or truth of the record of flight. The conditions of continuity and of freedom from knots must be respected and also a new condition peculiar to records involving the time dimension, namely, that no part of the curve of flight shall be so distorted that its slope represents a velocity greater than 3×10^{10} cms. per second. This last condition has to be added because velocities greater than that of light are not known to occur in nature.

It should be noticed also that the introduction of the time element removes the possibility of reducing the curve of flight to a mere point, the limit of reduction being a line; for although, to an observer sitting upon the flying bullet, the space-curve would be a mere point—the bullet having no motion in space relative to this observer—the progress of the bullet in "time" would still go on, the curve of flight becoming a line parallel to the axis of time.

If, to go a step further, we try to complete the record of the bat's progress in the same way that we have completed that of the bullet's progress, an awkward difficulty occurs. The bat's space record has already used up all the dimensions that are available for a chart, and no spatial direction remains in which to extend our block to provide for time. We require a 4-dimensional chart and no such chart can be made. There is, however, a simple solution of the difficulty, if we can be satisfied with a symbolical instead of a graphical method. The graphical method is the favourite method of the engineer when he wants to record some simple event such as, say, the change of the pressure (p) with the volume (v) of the contents of a cylinder. The mathematician usually favours the alternative symbolical method in which the equation to the curve takes the place of the graphic outline on squared paper. For example:—A given mass of air, held at constant temperature, is enclosed within a cylinder having an air-tight piston. Movement of the piston is accompanied by definite changes in the two factors pressure and volume, the latter decreasing when the former increases and *vice versa*. The engineer records this phenomenon upon squared paper, the vertical scale representing pressure and the horizontal scale (parallel to the bottom of the chart) representing volume.

Symbolical records

Having measured a large number of pairs of simultaneous values of p and v he marks by a dot the point on the chart corresponding to each pair. As a rule these dots, if numerous enough, suggest to the eye a fairly definite path: this the engineer draws in with his pencil as a continuous curve. In the case of the p-v curve he will find, and, in other cases, generally does find, that all the points upon the curve, not merely the actual experimental points, represent practical pairs of values of p and v. The engineer thus applies the law of probability when he draws in his continuous curve, for he says, in effect, that it is very unlikely that the random pairs of values chosen for measurement should fall upon a definite curve, unless, in fact, that curve represented the law underlying the phenomenon.

The mathematician works differently; he examines the actual experimental results and finds what relation of the "co-ordinates" p and v (though he prefers to call them y and x or else x_2 and x_1 no matter what sort of quantities they represent) are common to all the pairs of values given by the experiments: in other words he discovers what general relation there is common to all the particular pairs of values. He finds, for instance, in the case considered, that for all pairs of values of p and v, the product of p and v is constant, and he says, therefore, that the equation to the curve is, $p \times v = k$, where k is the constant whose actual numerical value results from the measurements. This equation contains exactly the same information as the graphical curve drawn by the engineer.

A 3-dimensional curve, such as that representing the vertical height (y), the horizontal distance (x) and the time (t), of a bullet's flight would be expressed in equations containing the quantities x, y and t and appropriate numerals obtained by experiment.

In general, for cases in which not more than three interdependent variable quantities are concerned, the two methods, the graphical and the symbolical, always are available at choice, but in cases, such for instance as that of the complete space-time record of the bat's flight, only the symbolical method can be used.

To work with more than three variables (even though they be called "dimensions") does not worry the mathematician, for he cares less about visualizing things than does the engineer.

It is impossible to form a mental picture of a 4-dimensional "block of glass" or of the curved "line" in it which would represent the complete space-time record of the bat's flight—it seems, in fact, impossible to form any "physical conception" of the "point" in a 4-dimensional continuum—but the mathematical expression of such a curve presents no difficulty, each point on the curve being represented by a set of values of x, y, z and t, which are called, respectively, the 3-space co-ordinates and the time co-ordinate of the point. Sometimes instead of x, y, z and t the symbols x_1, x_2, x_3 and x_4 are used.

World-lines

Without pretending to be able to visualize a 4-dimensional block, it is convenient to speak of such a thing, leaving to the mathematician the problem of putting our meaning into precise quantitative form. We may speak, then, of a 4-dimensional block of jelly in which we suppose the curve of space-time progress of a bat to be embedded. The space-time path of a mere particle is called the "world-line" of the particle—a term invented by Minkowski.

The point of view taken by Relativity, from which it is seen that none of the 3-space co-ordinates nor yet the time coordinate of a "point" on a world-line is "real," shows us that a world-line has no absolute shape—or that the term "shape" is without meaning in regard to a world-line. This is equivalent to saying that the description of a particle's progress in space-time, given completely by the world-line, is unaffected by any distortions which the 4-dimensional jelly, in which the line is thought of as embedded, may undergo.

The mollusc

In other words the world in which a particle exists may be described, legitimately, as a 4-dimensional "mollusc," and in order to remove the ambiguity which has been found unavoidable when physical phenomena are described in terms of the relative "motion" of particles of matter in space, the world of events may be thought of as a "static" structure of innumerable world-lines embedded in a mollusc-like frame of space-time.

We go back to the point at which the idea of the jelly or mollusc was introduced by reference to a 2-dimensional curve reproduced on a photographic film—a film which we supposed could withstand any amount of distortion. It was pointed out that distortion of the film had no "real" effect on the completeness or accuracy of the curve, since only the "shape" of the curve was changed and shape is not a real property, differently moving observers, with their different but equally warranted notions as to what things are "fixed," forming different but equally warranted judgments of the shape of a given line. The important fact was that even though the process of distortion had somehow obliterated the numerals on the scales of the chart the record still would be as complete and as accurate as ever! The curve was, in fact, meaningless to begin with and is still meaningless! The Relativist is one who uses the language of the universal observer, and to this observer space is unreal and the space path of a particle is without meaning. Meaningless also is the space-time path ("world-line") of the bullet as well as that of the bat.

Are we then reducing everything to nothing? If such phenomena as flying bullets and wheeling bats are to be explained away what then can we retain?

A moment's thought will show that the paths spoken of in the two cases—and considered as mere curved lines—were not, strictly, the paths of such complex things as a whole bullet or a whole bat, but merely were the paths of the centres of gravity of these bodies. The bullet and the bat acted only in their capacity as moving particles. A particle is the ultimate indefinable thing which we saw must be without meaning. All that an observer observes is the *relations* of particles in space-time, not the particles themselves.

Can a particle be observed?

The path of an *isolated* particle is unobservable, and the particle itself is unobservable. This means that an isolated world-line is unobservable—we know nothing *directly* about a world-line.

A particle exists in space and time—it *is* a world-line. The relations of particles, that is to say of world-lines, are all that the universal observer observes. What sort of observable relations can world-lines have? They are lines embedded in a jelly, and their observable relations are unaffected by arbitrary distortion of the jelly. Such quantities as the positions of world-lines and their shapes are unobservable—which, in physical science is equivalent to saying these quantities are unreal or non-existent.

What characteristics remain? There remain only the mutual contacts and intersections of world-lines and the space-time "ordering" of these. We can visualize the contacts and intersections of threads in 2- or in 3-dimensional space as well as express them by equations. In 4-dimensional space-time visualization is impossible; we can visualize neither a "world-line" nor even a "point" in a 4-dimensional continuum; we have to be content with the mathematical equations. For the visualization we fall back upon the artificial special case of a world-line in the 3-dimensional continuum of 2-dimensional space, and time.

Relativity says that the intersections of world-lines and the space-time "ordering" of these intersections are the observable realities of our world of events. Crossings (intersections) of world-lines are, in ordinary language, *collisions of particles*. Such collisions, or, as he would call them, space-time *coincidences* and their "ordering," are all that the universal observer regards as facts.

The reluctance which we, as individual souls, might feel in attempting to accept for the world of our own particular personal experience, a description in terms of space-time coincidences, is irrelevant to the acceptance of this description in the limited sense of science, namely the sense that it is the only description in which the universal observer, with his absurdly dogmatic notions about the equality of all observers and his weird language adapted to his ideas of "reality," is able to express the "facts" of nature.

Man as soul and man as observer

The distinction here drawn between a man as an individual human being—in which capacity he is probably always a solipsist—and the same man as a scientific observer, is one which we all make in practice, even though, explicitly, we may never have used the idea of the universal observer nor ever have sought in space-time coincidences the ultimate factors in observations.

Space-time coincidences

In order to be definite and unambiguous we try to record observations in terms of exact measurements. Every measuring instrument has some sort of a scale—as for instance, the scale of inches on a two-foot rule. Many have also a pointer—a clock, for instance, having a compound pointer consisting of a positively coupled pair of hands, and a scale of 60 minutes. In all cases the making of a measurement (or, as it is often called, the "taking of a reading") consists in recording a coincidence.

In using a two-foot rule to measure, say, the distance between two nails in a board we make the zero mark of the rule coincide with one nail and observe with what other mark on the rule the other nail coincides, and to avoid error due to the slipping of the zero mark while the other nail is under observation we have to make it a condition that the coincidences with the two nails shall be simultaneous, that is to say we look not merely for a space coincidence but a space-time coincidence.

In the case, also, of an instrument having a pointer which traverses a scale—as, for instance, the seconds dial of a watch—to make a measurement is to observe a space-time coincidence not a mere space coincidence. This becomes obvious as soon as we realize that a mere space coincidence is, if not without meaning, at any rate indefinable. Two points are coincident in space if they occupy both the *same position* in space—whatever that may mean. You and I are coincident if we both occupy the same chair. Need we add "at the same time"? Not if we are satisfied that the chair's position is fixed in space. The fact that we all realize that "fixed in space" is meaningless, accounts, perhaps, for the fact that "coincidence" never is used in any other sense than that of space-time coincidence. For aught we can know the scale division "60" on the seconds dial may possibly be at the same spot in space which the seconds hand will occupy when this stands at, say, "15"—the determining factor being the motion of the watch itself in "absolute space" (an expression which Relativity regards as meaningless). Of this, however, we do feel sure, that the "collision" of the seconds hand with a division on the dial, or indeed the collision of any two particles of matter is a space-time coincidence—a coming together both in space and in time.

Returning for a moment to the experiment of finding the distance between two nails we may now see that the measurement was incompletely described as *a* space-time coincidence. Strictly it was the coincidence in time of two space-time coincidences, namely, it was the time coincidence of the space-time coincidence at the zero of the scale and the space-time coincidence at the other observed division of the scale. And if now we admit that "coincidence in time" is meaningless, we concede one of the underlying principles of Relativity, namely, that the "distance" between two nails is not a real quantity at all, but is a relation between the system observed (two nails in a board) and the observer.

We see, however, that to take a reading on an instrument must always be to observe a space-time coincidence—the collision (or at any rate the closeness of approach which is regarded as equivalent to contact or collision) of the particle of matter whose part is played by the tip of a pointer, with another particle whose part is played by a mark upon a scale. The collision of two particles means the intersection or contact of two world-lines. Hence to state, for instance, that "the current is five amperes" is nothing more than to record the intersection of two world-lines. The relation of this fact with other facts constitutes the electrical phenomenon which is being examined, and as each factor of each fact is defined by a measurement and each measurement is the observation of an intersection of two world-lines, we begin to see that the whole of scientific experience can be described in terms of the intersections of world-lines and the space-time relations (that is to say the "ordering") of such intersections.

Some observations form an apparent exception to the statement that measurement consists essentially in the observation of coincidences. For instance, the chemist observes whether a precipitate is formed and what is its colour, or whether a substance is inflammable or whether it stains a borax bead. Such observations, however, are not measurements, they rather are indications serving to suggest to the observer what actual measurements he ought to make and what may be the range of magnitude of the quantities he will have to measure, such, for instance, as vapour density, specific heat, electrical conductivity, or refractive index. It is then the measurement of these quantities which gives the chemist his definite scientific facts, and every such measurement consists in the observation of coincidences.

The universal observer's point of view is not novel

Such considerations show that the point of view of the universal observer is not altogether novel to any man and is familiar to all who are accustomed to making measurements.

The universal observer's opinion about the world is that it can be truly and completely described in terms of the number and ordering of the intersections of world-lines, these world-lines being embedded in a 4-dimensional continuous jelly of space-time.

The intersection of world-lines

The Relativist contends that the distortion of the 4-dimensional chart of jelly, along with all the world-lines embedded in it, leaves the chart unchanged in regard to the "truth" of the picture it presents of scientific reality.

The meaning of this may be illustrated by reference to a 3-dimensional or even to a 2-dimensional space-time chart, for such a chart can be visualized. We suppose that a billiard ball is travelling along a 1-dimensional rectilinear course, with ever decreasing velocity and that its space-time progress is recorded on a sheet of squared paper, the time being measured along an evenly divided vertical scale and the distance along an evenly divided horizontal scale.

In principle the process of finding the data for the record may consist of marking out along the path on the table, which the ball will follow, a scale of inches, and noting by aid of a clock the exact instant at which the ball's position coincides with successive inch divisions of the path. When the Relativist says that the space-time curve thus obtained may now be distorted without effect upon the record, he means that it is in reality immaterial to the "truth" of the record how the "inch" divisions be spaced along the path or how erratically the clock may be going. The path divisions may be a random series of spaces of all sorts of odd lengths and the time divisions also may be random.

Since, however, such a chart contains only one world-line and no intersections, it is, according to the universal observer, a meaningless record and we need not be surprised that the actual spacing of the so-called "equal" divisions, whether of space or of time, is immaterial.

In order to introduce "reality" (i.e. intersections of world-lines) into the phenomenon we suppose that a series of balls is sent along the same path one after another and that each ball has a greater velocity than its predecessor and so overtakes it. In practice the balls on colliding would rebound, but to simplify the illustration we will suppose that when a ball overtakes another it simply rolls over it, neither ball suffering any disturbance in velocity. Then the overtaking becomes a simple intersection of two world-lines. Upon an evenly divided undistorted chart we might then get a series of world-lines such as *A*, *B* and *C* in Fig. 2, giving the intersections *P*, *Q* and *R*. These points *P*, *Q*, *R* are space-time coincidences. They are realities, for however the diagram may be distorted these intersections must always remain—like knots in a net—and not only will the intersections *P*, *Q*, *R* remain but something else also, namely, their space-time ordering which also is "real." The world-line *A* begins at *O* and consists of a continuous series of points among which the points *P* and *Q* occur, and no matter how the chart may have been distorted the order *OPQ* will always persist; it can never become *OQP*.

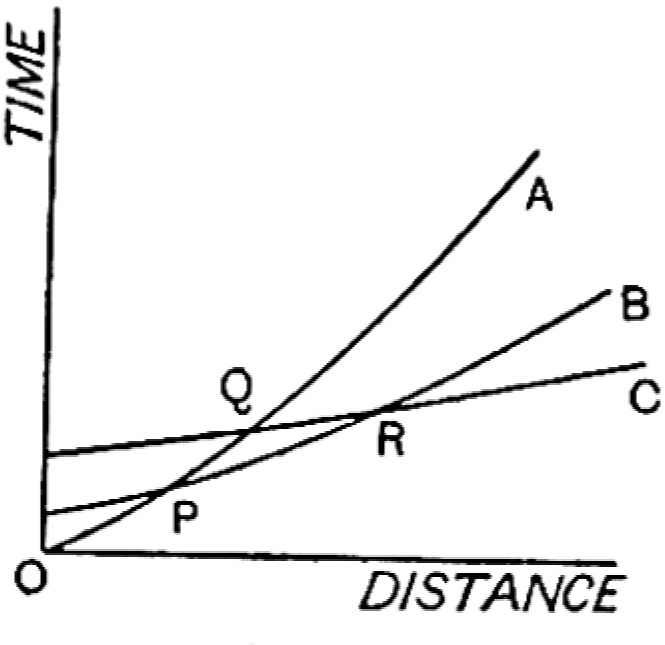

FIG. 2

It will be remembered that the impossibility of knowing in what "absolute" manner the linear dimensions of bodies may be varying and the going of clocks may be changing, underlies the Relativist's contention that space intervals and time intervals are unreal and that only the intersections of world-lines are similarly observed by all observers. We have now to ask, what is the assumed property of space intervals and time intervals which justifies the Relativist in attributing reality also to the space-time *ordering* of intersections? The assumptions are these: That whatever erratic changes may happen to space intervals they never become less than nought—that is to say whatever changes may occur in the length of a rod carried by an observer that length never becomes less than nought—and, in regard to time intervals these also can never become negative —that is to say however queerly an observer's clock may go the beats of its pendulum can never take less than no time.

These conditions put restrictions upon the distortion of charts.

The problem before the Relativist

Using his own peculiar notions of space and time and his own standard yard rod and chronometer, an individual observer, *A*, makes his own precise record of a phenomenon, and this record seems to him to be the only possible *true* record. The scales of his chart are evenly divided straight lines, and each world-line he inserts has its own definite shape and position relatively to the chosen rectilinear frame of reference. As we have seen, the bases of the construction of this chart are this particular observer's notion of what things are "fixed" in the ether and his contention that the ether is a rigid, uniform, distortionless, homogeneous medium pervading the universe.

Another observer, *B*, moving relatively to *A*, attributes fixed position in the ether to some other set of bodies, and makes of the very same observed system a different world chart. He follows the same rules which *A* followed about making measurements and about drawing accurate charts, being careful to use properly standardized scales and clocks and to draw "truly" rectilinear and mutually perpendicular axes of reference with uniformly divided scales. Nevertheless, since *B*'s notion of what is fixed differs from *A*'s, it is probable that *B*'s notion of what is straight will also differ, and in any case it is certain that his notion of direction in space-time will differ from *A*'s (compare the path of the chalk in reference to the lift well with the same path in reference to the wall of the falling lift—p. 79).

Other observers draw still other, different, "true" charts, but to all these charts two things are common, namely, the number and the ordering of the world-line intersections.

One of the first problems of the Relativist who wishes to justify his position is to express what it is precisely that he means by the contention that the intersections of world-lines and the space-time ordering of intersections are common factors, and indeed the only common factors, in the varying descriptions made by different observers.

We have seen already that the differences between descriptions result from the use of different frames of reference. This has to be expressed mathematically and, in addition, it has to be shown how the measurements made by individual observers —for such measurements are all that science has to work with— can be made to disclose properties of the world which shall be found by *all* observers. Something must first be said, therefore, as to the way in which mathematics can be applied to science, in particular that branch of mathematics which is required in an investigation of the agreement of experimental facts with the Theory of Relativity.

The mathematical method

A theory can be tested only by experiment, and as quantitative measurements expressed by numerals are the only intelligible results that experiments give, the theory that is to be tested must be expressed in a form to which numerical quantities can be applied. For instance, by making many experiments with falling bodies I am led to form the theory that the vertical distance in feet through which a body falls from rest must be equal *in all cases* to 16·1 times the square of the number of seconds occupied by the fall.

I arrived at this theory by making a score of experiments in each of which the body was allowed to fall some different measured distance, each fall occupying a measured number of seconds. I plotted these results upon a chart and found that all the points fell approximately upon a smooth curve. This led me to suppose that if I had made other experiments, these also would have given points falling on the same curve. The theory I thus formed was that every one of the infinite number of pairs of numerical values of the space "*s*" and the time "*t*" which could be read off from the smooth curve would agree with the corresponding experiment if it were made.

There are available three ways of expressing my theory:

1st. The verbal expression: When a body is allowed to fall freely from rest it covers a vertical distance in feet equal to sixteen point one times the square of the time of fall in seconds.

2nd. The symbolical expression:

$$s = 16{\cdot}1\, t^2.$$

3rd. Expression in the form of an unending series of numerical statements, as, for instance:

1 ft.	will take	$(1/\sqrt{16{\cdot}1})$	second,
4	,,	$(2/\sqrt{16{\cdot}1})$	,,
9	,,	$(3/\sqrt{16{\cdot}1})$	,,
16	,,	$(4/\sqrt{16{\cdot}1})$	,,

and so on.

The second method commends itself for convenience and clearness: it is the mathematical method.

In the simple case just given the symbols "*s*" and "*t*" are called the "variables," and it is understood that if *any* desired value be assigned to "*s*" or to "*t*" the solution of the equation will give the corresponding value of "*t*" or of "*s*." The equation expresses the general law underlying all possible particular cases.

The mathematical method uses symbols not only to represent measurable quantities but also to represent operations.

For instance, the rate of change of a quantity "y" in respect of "x" is written (dy/dx). If "y" is the distance of a moving particle from its starting point and "x" is the time elapsed since starting then (dy/dx) is the rate of change of distance in respect of time, that is to say it is the velocity with which the particle is moving relatively to the starting point. The symbol "d" means "a very small change of" and the expression dy/dx represents the operation of imparting to the particle a small change in "x" and dividing this change into the consequent small change thus produced in "y."

The data which scientists give to the mathematician for investigation are measurements made with ordinary scales and clocks. Relativity assumes that the frame of reference, which the observer believed to be rectilinear and fixed, may actually have been distorted and in motion. In other words—from the legitimate alternative point of view—Relativity assumes that the observer's scales and clocks may have been "wrong"—possibly with a continuously and erratically changing wrongness—but that no criterion is known as to the amount of the wrongness. The observer is not directly aware of any error in his carefully calibrated measuring devices, but when this error is due to accelerated motion of his frame the observer unwittingly discloses this fact to the Relativist by including "forces" in his data.

A ball thrown across a freely falling lift, by a man in the lift, describes a straight line within the lift. The man in the lift describes the phenomenon as straight line motion and says no gravitational force is present in the lift. A man on the ground outside the lift finds that the ball describes a parabolic path (relative to the earth) and he describes the phenomenon as that of parabolic motion under the influence of the "force" of gravitation.

The Relativist tries to use always a frame of reference so chosen that no forces appear, and when an observer presents data in which forces do appear he sets this down to the use by the observer of a distorted frame of reference; he knows, for instance, that a line which the observer calls parabolic might just as truly have been called straight and that the "rectilinear" axes which the observer believes he used might just as truly have been called parabolic or otherwise curved.

Difficulty of avoiding confusion in the discussion of Relativity

At this point it is appropriate to emphasize the unavoidability of a tendency to confusion when the application of Relativity to practice is under discussion. Practice means practical quantitative measurement, the data of which are *lengths*, *times*, *forces*, *masses*, etc. Relativity is a method of description from which length, time, velocity, acceleration, mass and force, have been eliminated.

Confusion arises also through the use of the two alternative points of view in regard to the "cause" of the differences between the descriptions of an event as given by differently moving observers, namely the point of view that motion through the ether makes all bodies get shorter and all clocks go more slowly, and the alternative point of view that a moving observer uses a distorted frame of reference. Neither of these viewpoints is strictly correct, that is to say neither is recognized by the universal observer, who knows nothing of lengths or times and nothing of their changes, nor appreciates the notion of shape or of distortion. These points of view are useful for helping the individual observer to feel on more familiar ground while trying to follow the strange paths of the Relativist, but they must be used with care if error is to be avoided, and ultimately they must be given up.

Strictly—that is to say from the Relativist's viewpoint—the attempt to explain phenomena in terms either of the shortening of moving rods and the slowing of clocks, or in terms of the distortion of frames of reference, is a reversion to the Newtonian ether idea and is an evasion of the Theory of Relativity—the only theory consistent with all known phenomena. Length and time intervals are unreal, and so are shape and distortion. A world-line has a "path" in 4-dimensional space-time, but shape is not a property which can be attributed in any definable way to a world-line. A world-line is composed of "intervals," but these are space-time intervals having a peculiar mathematical definition, they are not length intervals.

The Theory of Relativity concerns the "real" world of physical science, and as this happens to be a world which, as yet, no individual can visualize or picture or "physically conceive," no "physical explanation" of the theory is at present possible, but only a symbolical mathematical one, and it is futile for anyone unacquainted with the mathematical method to hope to learn anything definite about Relativity.

"Forces" are the Relativist's clue

The important clue which enables the Relativist to investigate practical matters is "force." The ordinary observer introduces "forces" whenever his point of view as to what is "fixed" begins to make him attribute "unnatural" shapes to the paths of particles. The ordinary observer says that bodies move in *straight* lines unless "forces" prevent this; the Relativist says that bodies move always in their natural paths ("geodesics").

The ordinary observer defines a "straight" line as the path taken by a body which is acted on by no "forces"; he then defines "force" as that which deflects a body's motion away from a "straight" line. The Relativist says a "geodesic line" (or, simply, a "geodesic") is the path which a particle follows in space-time—he tries no further to define such a line.

Thus the Relativist postulates an indefinable mystery, while the ordinary observer side-steps the mystery by a circular argument. Obviously the mystery still exists; the Relativist asks where the advantage can be in having camouflaged it by a circular argument.

Circular arguments

The objection to every circular argument is that it camouflages ignorance, covering, when the ignorance is universal, a mystery, when it is particular to the man who speaks, an error. Anyone who insists on speaking of "forces" may, if he likes, regard a "geodesic" as a path along which a particle is able to avoid all forces.

The method of Relativity

The method used by the Relativist when he wishes to give simple explanations has been illustrated by the experiment of the ball thrown across a falling lift. His method is to discover what sort of motion imparted to the lift will convert the lift walls into a frame in respect to which the force of gravitation no longer exists. This particular case seems simple but we shall find that the general problem is not adequately illustrated by such a simple case.

In thinking of the lift, attention was confined to the small local region of the lift and its well. For this small region experiment showed that to choose an accelerated frame of reference (namely that of the freely falling lift) sufficed to get rid of the force of gravitation, or explain it away. This suggests that gravitation is not real or absolute—for the absolute is that which appears to all possible observers; it is something which no observer, however his situation or motion be chosen, can fail to observe, that is to say it is something which cannot be explained away.

We shall see, however, that gravitation cannot be explained away. Gravitation is the expression of something absolute in nature, and it is as true as it sounds strange to say that the demand for the General Theory of *Relativity* has arisen out of the obstinacy with which gravitation retains its *absolute* nature.

CHAPTER V

Gravitation cannot be explained away

EVERY gravitational field extends throughout the universe, and to explain away the earth's field we must explain it away everywhere, not merely in one small region. Actually the elimination of gravitation even from the small region of the lift was incomplete. The acceleration of a lift removes the gravitational force completely only from those particles of the lift whose motion is directed precisely along the direction of the gravitational force, and since the motion of the particles constituting the rigid lift is in strictly parallel lines, while the force of gravity, acting towards the earth's centre, is in converging lines, it follows that only those particles lying in the vertical axis of the lift (the vertical centre line of the lift) can be freed entirely from gravitation.

Thus the path of a ball thrown horizontally across a stationary lift is not strictly a parabola in regard to the frame formed by the lift, it is the combination of a parabolic curve with the curvature of the earth. When the lift as a whole is falling freely its outer walls are being accelerated (towards the earth's centre) a little less than is the middle of its floor and so the motion of these walls does not quite compensate the bend in the ball's path, and this path, referred to the freely falling lift, is not strictly straight but is just a little concave towards the earth. The larger the floor area of the lift the less perfect is the compensation.

To make the difficulty still more apparent let us imagine that the lift is mounted on a perfectly lubricated steel rod which forms the vertical axis of the lift and penetrates the earth diametrically from here to the antipodes. We put our observer first of all in the lift which is on this side of the earth, and as the lift is falling freely he finds that gravitation has been eliminated, his hat, for instance, exerting no force upon his head. We next put him—the lift and its rod still moving freely—on the rod at the antipodes (where again the rod stands out vertically) and there he finds that the rod's motion is vertically *upwards* with an acceleration equal and opposite to that of gravitation, and his hat now presses with double its proper weight upon his head.

Thus although our choice of frame of reference (the frame of the freely falling lift) has succeeded in eliminating gravitation inside the lift at this end of the earth's diameter, its effect at the antipodes has been not to eliminate but to double the force of gravitation.

Centrifugal force

The earth's centrifugal force presents no such difficulty, for this force can be explained here and at the antipodes and everywhere simultaneously, by simply choosing a frame rotating with the earth. This point must be made clear. We consider a rotating disc of metal. Knowing its radius and its speed of rotation we find the peripheral velocity at any point on the disc. Calling this velocity V we know that any body of mass M fixed on the disc at this point will experience a centrifugal force equal to (MV^2/R) where R is the radius of the periphery considered. If the disc is large enough I may sit upon it and make observations and I know that I shall have to stick or tie myself tightly to the disc "to prevent being thrown off by the centrifugal force," or, to express the matter from another point of view which is equally legitimate, "to prevent my remaining *where I am* while the disc slides beneath me."

Centrifugal force can be explained away

After I have fixed myself to the rotating disc it occurs to me to suppose that the notion held by me when I was an outside observer, namely that the disc was rotating, may have been unwarranted. I am now on the disc and I perceive no rotation—all that I perceive is a force tending to throw me away from the centre. Why should I not call this a new sort of gravitational force? It is a force which I find gives equal accelerations to all bodies when these are let loose, and this is exactly the most characteristic of the properties which ordinary gravitation possesses. Moreover it can be explained away locally (e.g. inside a liberated box) just as gravitation can.

There is, however, one outstanding peculiarity about centrifugal force which seems to put it outside the class of forces called gravitational. Centrifugal force can be explained away everywhere at once, not merely locally.

If the disc were very slippery a man sitting upon it would simply remain "where he was," while the disc would slide under him—he would experience no centrifugal force. Similarly another disc at rest relatively to the rotating one might be laid down upon the latter and would experience no force anywhere. Being frictionless the rotating disc would go on moving beneath the idle disc and in respect to the frame of reference represented by this idle disc there would be no force anywhere.

A Japanese juggler uses this principle when he causes a coin to course around the edge of a rotating parasol. The coin experiences no centrifugal force.

If I lived on a rotating disc all I should have to do to get rid of centrifugal force would be what the juggler's coin does, namely, "remain at rest," that is to say move circumferentially in the proper direction and at the proper speed around the disc. In a motor car thus moving I should detect no centrifugal force, and if the whole disc were covered with motor cars moving in concentric circles about the centre of the disc at appropriate speeds, not one of these cars would experience centrifugal force. The appropriate speeds would be such that all the cars remained at rest relatively to one another—in other words the cars might be regarded as coupled together to form a rigid frame of reference.

We conclude that there is the distinguishing difference between ordinary gravitational force and centrifugal force that the former cannot but *the latter can be explained away everywhere at once*.

Consistently with our statement about gravitational force that since it cannot be explained away it ought to be regarded as the expression of some absolute factor in nature, the conclusion seems now to be warranted that as centrifugal force *can* be explained away it cannot be the expression of anything absolute such, for instance, as "absolute rotation," and that in fact the Newtonian conception of absolute rotation is unwarranted. We shall see, however, that no such conclusion is justified; its apparent justification being due to a misunderstanding of the term "explained away."

The meaning of "explained away"

We have seen that gravitational force as a whole cannot be explained away. It is not permissible, however, to conclude that gravitational *force* is real. No force is real, for force is a relative quantity; force is a relation of an observer to an event, it is not a factor in the event. The deduction which is permissible is this, that the existence of a field of gravitational force is the expression of some absolute factor in nature.

Centrifugal force as a whole can be explained away; it is permissible therefore to conclude that centrifugal force is not real—we know also on general grounds that no force is real—but it is not permissible to conclude that the thing which it characterizes, namely, rotation, cannot be real or absolute.

No relative quantity is real, nevertheless any such quantity may possibly be the expression of something absolute, for there can be no relation unless there be factors which bear the relation and one or more of these factors may possibly be real.

We shall return later to the question of the absolute nature of rotations: it is necessary now only to avoid the unwarranted conclusion that centrifugal forces cannot be an expression of factors in nature which are absolute.

A thing which "cannot be explained away" is one which has *some* effect on every observer—but not necessarily the same effect. For instance, the earth's gravitational field of force has some effect on every observer but different effects on different observers. The observer in the falling lift would perceive the gravitational effect at the antipodes—for, of course, the range of an observer's field of measurement is supposed to be unlimited, that is to say we postulate for each observer the power of extending his own chosen frame of reference throughout all space and making measurements in all parts of it. An observer in a stationary lift perceives a force locally and also an equal force at the antipodes.

Thus a thing which "cannot be explained away" is not necessarily real or absolute, for reality has been defined as a thing which has not merely *some* effect but which has the *same effect* on every observer.

A thing which can be explained away is one which some specifiable observer will be unable to observe, it certainly cannot be a real or absolute thing, it must be but relative; but it *may* express a relation to an observer, of things which themselves are absolute.

The practical data which the Relativist has to translate in terms of intersections and the ordering of intersections of world-lines, include fields of force; some being fields which can be explained away, others fields which cannot be explained away.

The problem of the reality underlying gravitation

In examining the phenomena of gravitation and of rotation we have found that no observer, however his frame of reference may be moving, could fail to observe gravitation, but that centrifugal force would be totally absent to an observer whose frame of reference rotated. We conclude that centrifugal force is unreal although possibly it may result from something which is real, but that something real *must* underlie gravitation.

The Relativist's first problem is to express the reality underlying gravitation.

It may help our explanation if we compare the imagined case of two similar discs A and B, the former rotating and thus possessing a field of centrifugal force (its own gravitational field being left out of account) and the latter (B) at rest and possessing only a gravitational field.

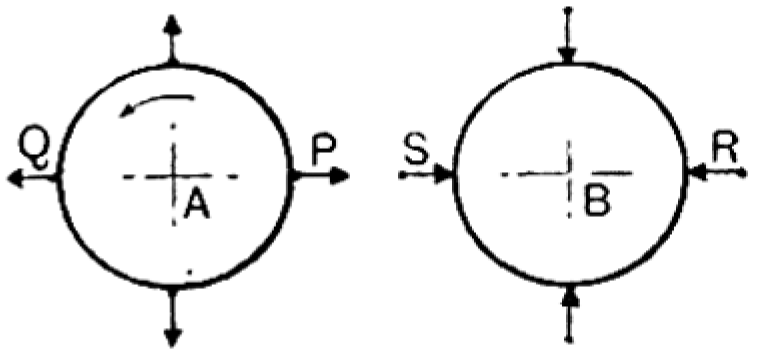

FIG. 3

The arrow heads in Fig. 3 indicate the paths which particles lying in the disc's plane would take (*relative to the discs*) at the instant they were made free, in the case of A, to leave the disc, in the case of B, to fall towards the disc. The four particles indicated by dots on the arrows have been taken as examples.

There are two ways of eliminating the centrifugal field; to eliminate it locally and to eliminate it totally. We imagine *P* to be equipped with a frame of reference represented by a cubical structure of straight lines, as in Fig. 4. *P* is at the origin of the frame and as *P* moves the frame moves with it. The disc itself we imagine to be extended so as to form a background for the motion of *P*.

At the instant of its liberation from the disc the particle *P* is moving vertically upwards; it describes, therefore, on a fixed background a vertical straight line.

In reference, however, to the extended disc *P* starts out radially and then sweeps downward as in Fig. 5. This is because the point on the disc at which *P* started out moves away to the left of *P* and so brings *P* into regions of the extended disc background in which the peripheral velocity of the background is ever becoming more and more in excess of *P*'s original vertical velocity.

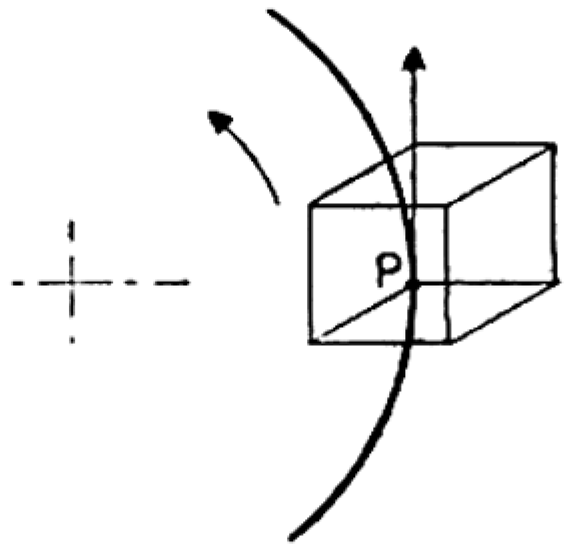

FIG. 4

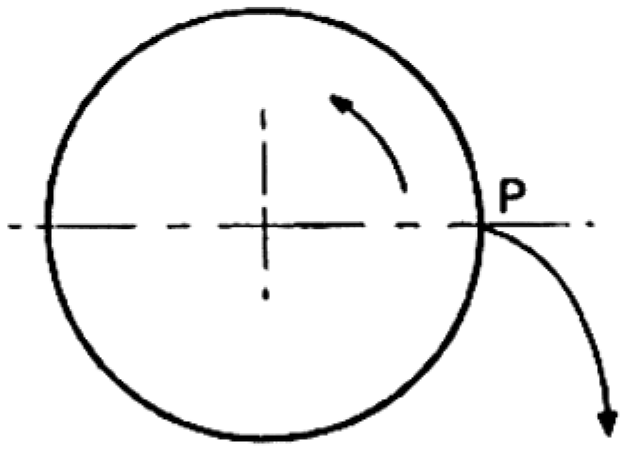

FIG. 5

The bend in the curve of Fig. 5 is due to centrifugal force. *P* is a free particle; a Newtonian observer on the disc finds that *P*'s path is not what he himself calls a straight line, hence he says there must be a "force" acting on *P* and "deflecting" it from the straight course; this force which he has invented he calls centrifugal force.

An observer at *P* in the cubical frame would experience no force for he remains at rest in this frame. Thus the force has been removed locally. That the removal is only local may be seen by considering that within the same frame of reference containing *P*, extended to embrace the opposite side of the disc, the particle *Q* will not be at rest but will be moving in a curved path and, therefore, experiencing force.

Thus in a frame of reference "falling" freely centrifugal force has been eliminated locally.

The frame just chosen was one in which *P* was at rest, but it is not necessary for a particle to be at rest in a frame in order that it should experience no force, it may be either at rest or in uniform rectilinear motion. In respect to a stationary background *P* is in rectilinear motion retaining uniformly the velocity *V* which it had on leaving the disc. Hence a stationary background is a frame in reference to which the force on *P* is eliminated. A stationary background is related to *Q* and other particles in the same way as to *P*, that is to say, in respect of such a background centrifugal force exists nowhere.

In the case of gravitation local elimination is all that can be effected.

A pair of discs was chosen for our illustration because, with these, the forces, whether centrifugal or gravitational, may be thought of as confined to a 2-dimensional plane (that of the discs) and the world-lines of the particles can, therefore, be visualized, the third spatial dimension—measured, say, downwards perpendicularly to the plane of the discs—serving for the development of the particle's path in "time."

On the extended disc as a background the world-lines of the particles *P* and *Q* before leaving the disc are a pair of vertical straight lines going down into the paper. After liberation of *P* and *Q* from the disc the world-lines are as indicated by the curves *P* and *Q* in Fig. 6, in

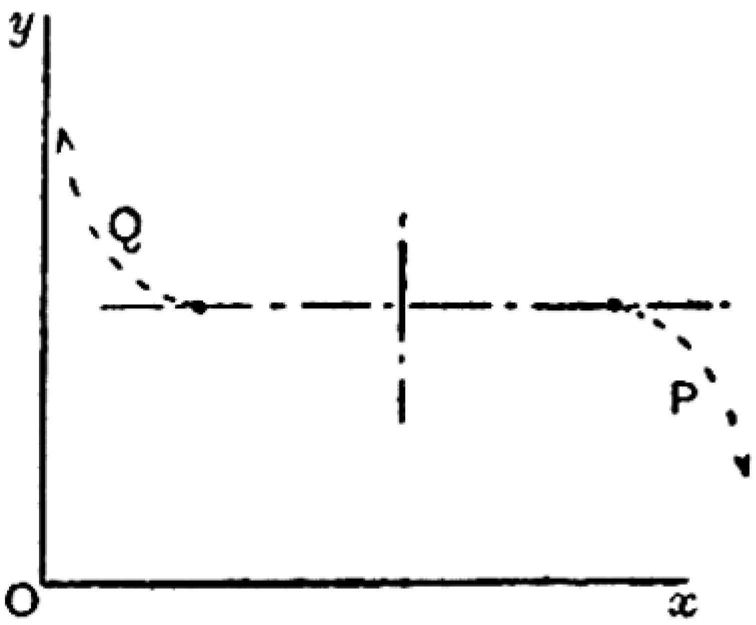

FIG. 6

which the wider spacing of the dots is intended to suggest the recession of the curves into "time" (beneath the plane of the paper).

The world-lines for corresponding particles *R* and *S*, falling under gravitation, are shown in Fig. 7.

Fig. 8 shows the world-lines of Fig. 7 seen from the direction *yO*.

These figures bring out the important conception that when an observer says a free particle is being acted on by a force he means that he considers the particle's world-line to be bent. Often—as for instance in Fig. 6 (centrifugal force)—the bend is "in space," but often also—as in Fig. 8 (gravitation)—it is "in time." The effect of gravitation upon the path of a bullet fired horizontally from the top of a cliff is both a spatial bend and a bend in time. As, usually, the path of such a bullet is in a 2-dimensional spatial plane we can visualize its world-line.

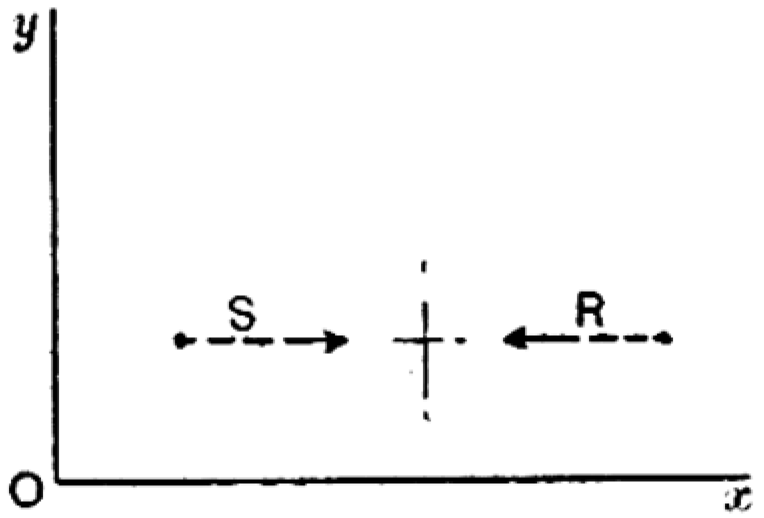

FIG. 7

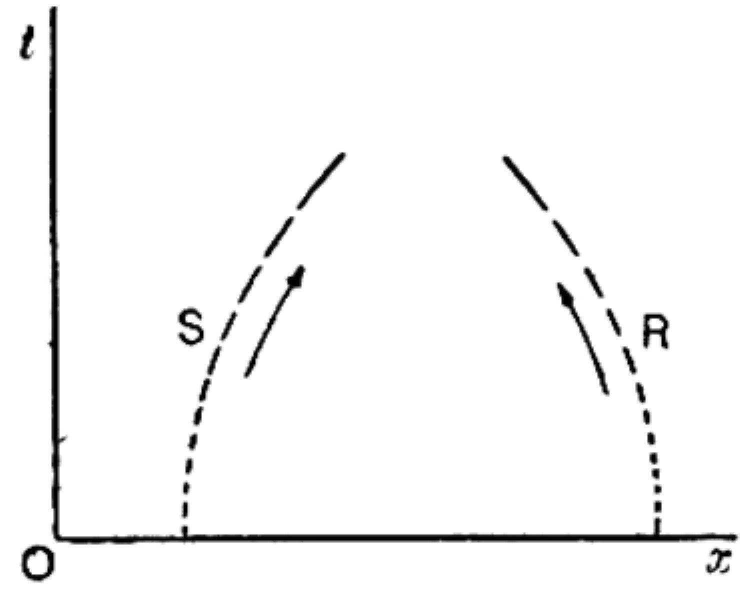

FIG. 8

In Fig. 9 the gun is at *P* on the vertical space-axis *y*, the *x* axis representing a horizontal base line.

In Fig. 10 we take a view of Fig. 9 along the direction *yO*, and so get a curve showing how distance travelled horizontally from the gun's mouth varies with time.

In practice the horizontal velocity will fall off as the *x* distance increases, because of air-friction, and the curve will bend as shown in Fig. 10. This bend is "due to" the force of air-friction. If there were no air the bullet's horizontal velocity would remain constant and the curve of Fig. 10 would be a straight line.

Fig. 11 is the view taken along *xO*, it shows the time bend in the bullet's world-line due to gravitational force. Thus *OT* is the total time of the bullet's flight, during which it falls the vertical distance *YO* (the height of the cliff). That the shape of the curve must be of the sort shown in the figure is obvious when we consider that the downward velocity of the bullet is constantly increasing on account of gravitational acceleration.

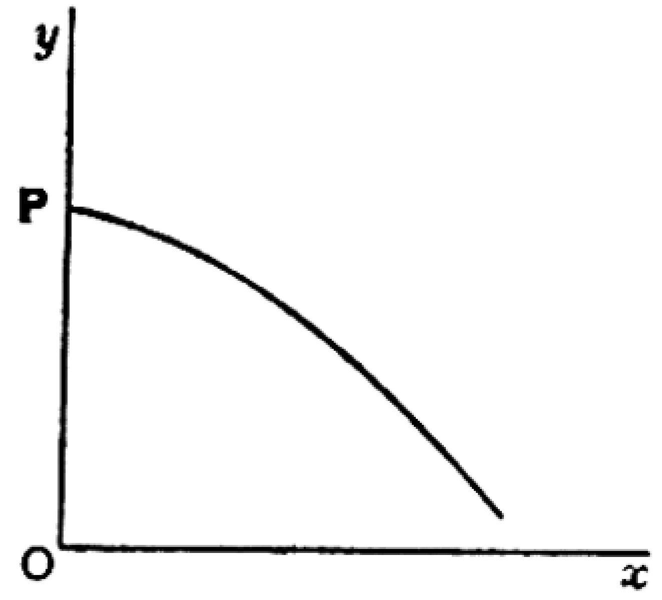

FIG. 9

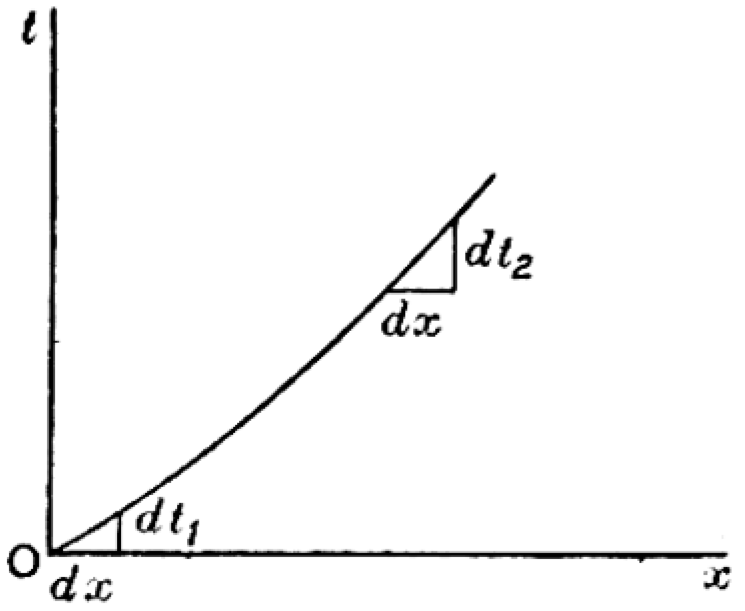

FIG. 10

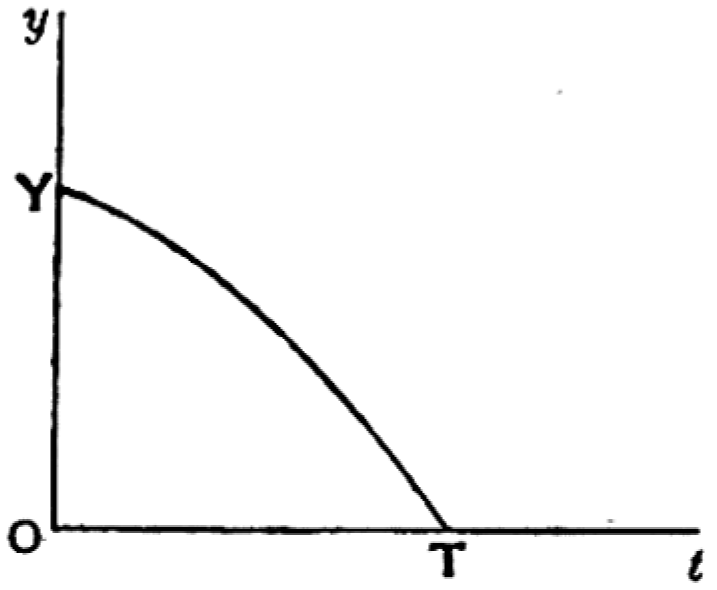

FIG. 11

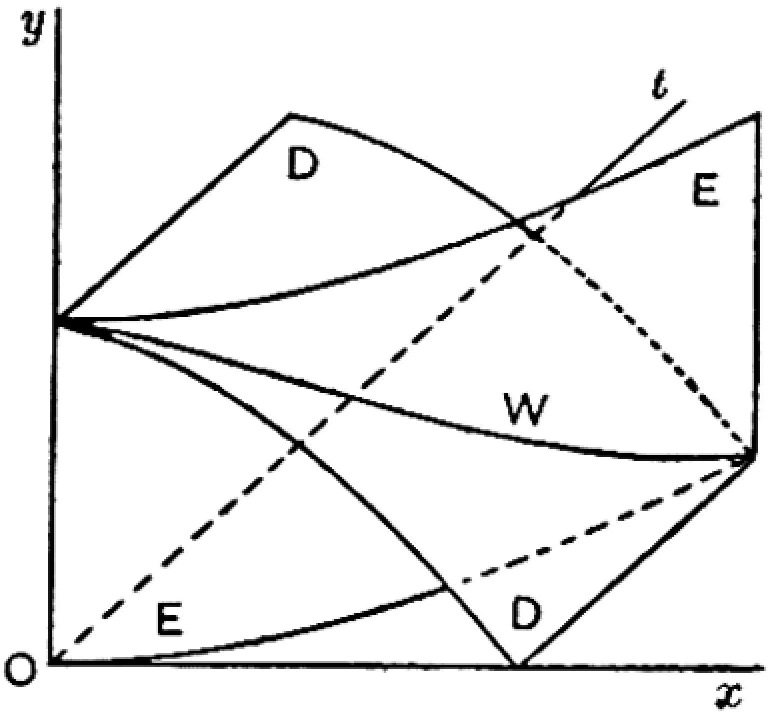

FIG. 12

It is not easy to indicate by a line drawing the 3-dimensional curve which completely represents the world-line, but an attempt has been made in Fig. 12.

In order to show how the world-line *W* is compounded of the curves of Figs. 9 and 10 the surfaces *DD* and *EE* have been shown whose mutual intersection is the world-line *W*. Looking at Fig. 12 along *yO* we see Fig. 10; looking along *Ot* we get Fig. 9.

The "straightening" of curves by choice of frame

We return now to Figs. 6, 7 and 8 (pp. 109, 110) and have to see as clearly as may be that by choice of a new frame of reference the curves *P* and *Q* of Fig. 6 may both at once be "straightened," while no such simultaneous straightening is possible for the curves *R* and *S* of Fig. 8.

In all three figures the *x* and *y* axes are mutually perpendicular lines fixed in relation to the discs—they might be scratches made on the imagined extensions of the discs—and the time axis, *t*, is a straight line perpendicular to the plane of the discs, drawn through the intersection, *O*, of the *x* and *y* axes (see Fig. 13). The curves *P* and *Q* are the traces which two diametrically opposite particles, rotating with the disc about its centre *C*, in a counter-clockwise sense, would make upon the extended disc.

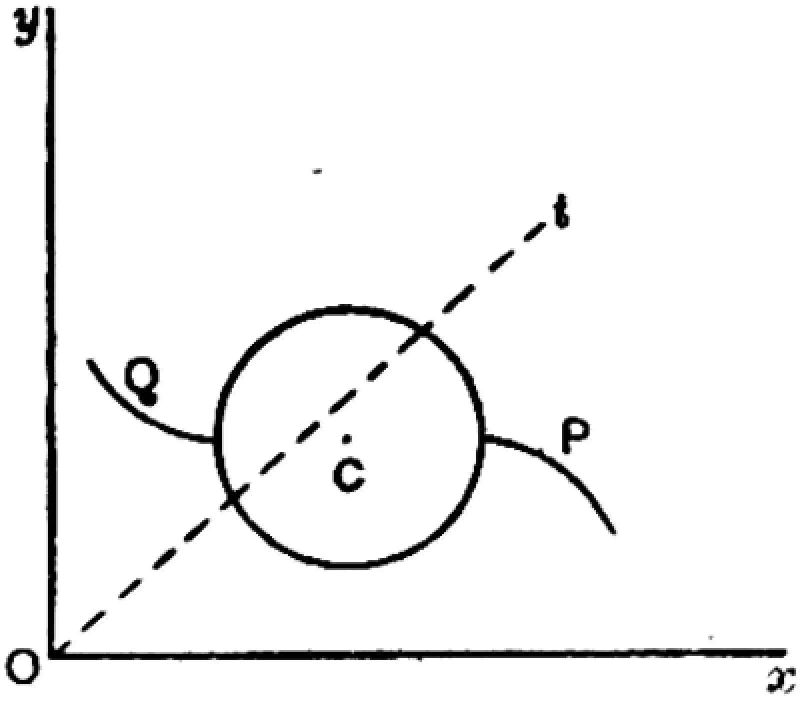

FIG. 13

In order to become familiar with these curves of centrifugal force it is worth while actually to plot them. The plotting is made easy by aid of a simple device. A disc of thick cardboard, say, six inches in diameter, is glued concentrically upon a disc of stout drawing paper, say, 18 inches in diameter. A pin through the centre *C* (Fig. 14) fixes the system to a drawing board and forms a centre about which the discs may turn.

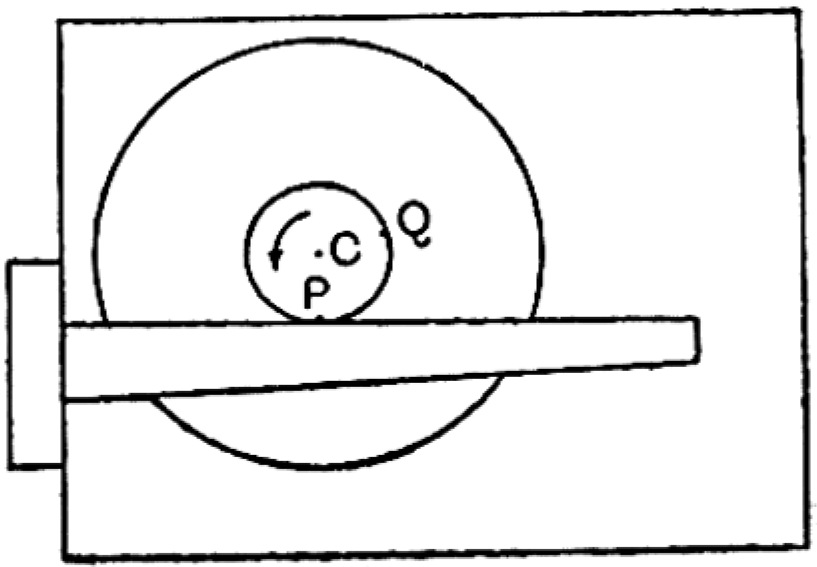

FIG. 14

A thread, having one end fixed, say at *Q*, is taken nearly around the cardboard disc and is held taut by means of the point of a pencil held upright at *P* where the edge of a T-square touches the disc. Holding the T-square fixed as shown in Fig. 14, the pencil *P* is drawn along the edge of the T-square, its point making a trace upon the disc of paper which, drawn by the thread, rotates beneath it. This trace represents the path which a liberated particle at *P* would take under the action of centrifugal force.

It will be noted that the edge of the T-square (a straight line tangential to the disc) is the path of the same particle, *P*, in the drawing-board frame of reference.

Comparing Fig. 6 with Fig. 8 we see why it is that while proper choice of a moving frame of reference may possibly straighten both *P* and *Q*, no conceivable choice of frame can straighten both *S* and *R*.

Both *P* and *Q* suggest clockwise rotation and it is to be expected that if a frame of reference be made to follow the same rotation the curves *P* and *Q* may prove to be straight lines in respect to this frame. The frame must rotate clockwise relatively to the disc from which the particles are being liberated.

S and *R* suggest oppositely directed rotations, hence no rotating frame can straighten both *S* and *R* at once. As, however, *S* and *R* both point "upwards" the suggestion may be worth considering that a frame of reference moving with properly chosen acceleration upwards might be found in respect to which the curves might prove to be "straight." But "upwards" in Fig. 8 means in the "time" direction and everything in the world always does move in the time direction (of each observer concerned) and we have defined "time" in such a way that the idea of altering the speed of a thing's passage in the time direction is meaningless—time flows uniformly in respect to all things in the Newtonian world, not excepting frames of reference. It must be remembered throughout this discussion, that when we deal with space and time separately we are adopting the point of view and the language of some particular Newtonian observer, not those of the universal observer.

Thus to accelerate the frame of Fig. 8 along the Ot direction is meaningless. There remains the suggestion of imparting motion along Ox, but since S bends in the direction O to x while R bends in the direction x to O, it is obvious that any motion imparted to tOx which will straighten one of the two curves could only add to the bend of the other; simultaneous straightening of both S and R by such a method being out of the question.

Our object was to specify some frame of reference in respect to which the accelerations of two falling particles, located one here and the other at the antipodes, might be simultaneously annihilated—just as, in the frame of a falling lift the acceleration of one particle can be annihilated. Thus far we have failed.

The attempt will have to be continued along new lines, but it is important first to become more familiar with the nature of the difficulties met in the attempts already outlined.

In taking a pair of particles each at the antipodes of the other we have seen the difficulty more clearly than it would have appeared in the case of particles near together on the earth. In principle the same difficulty still exists. It has been pointed out already that, strictly, the force of gravitation within a freely falling lift disappears in regard to the frame of reference formed by the rigid lift only for those particles which lie upon the vertical centre line of the lift, this line passing through the earth's centre while the walls of the lift—being parallel and not convergent—do not. Particles in the lift walls do not fall *freely*, for in falling freely a particle must fall along a radius of the earth.

Strictly, therefore, the curvature of the *y-t* curve for one falling particle in the centre of the lift is not quite like that for a neighbouring particle falling freely near one wall of the lift. The difference is almost immeasurably small but we are concerned with the principle underlying this difference, not with its magnitude.

The problem before us is to discover how an observer's rigid frame of reference must move in order to straighten, simultaneously, two curves whose curvature is unequal.

"Straightening" the paths of falling bodies

Let us suppose that *A* and *B* in Fig. 15 are the curves which have to be straightened. *A* might be for, say, a particle falling freely along the middle vertical line of a stationary lift, the line *Oy* being thought of as a string or tape measure stretched from floor to ceiling in the centre of the lift and *Ot* (the axis along which time is measured) as a straight line drawn from the foot of this string along the lift floor. Then the curve *A* will be a parabola. Curve *B* refers to a second particle falling freely from the same height but near to the side of the lift. In order to realize that *B* must be less bent than *A*, and slightly more like a straight line parallel to *Ot*, we regard for a moment the extreme case of an imagined lift whose frame embraces the whole earth (Fig. 16).

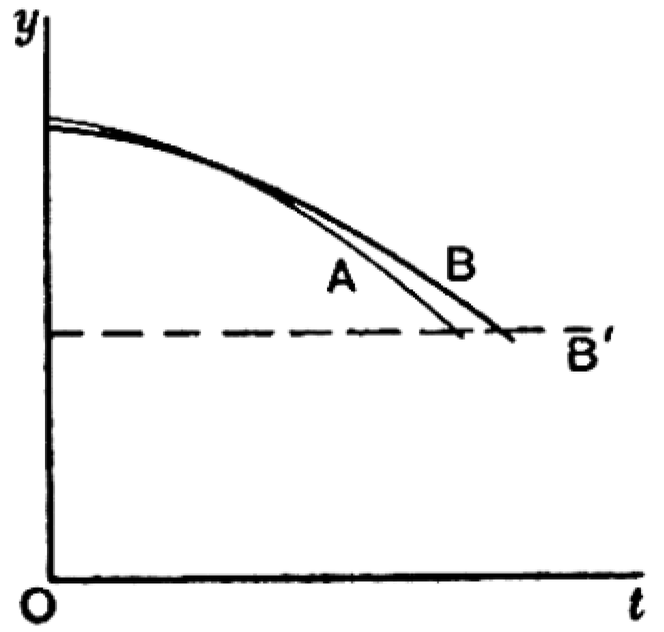

FIG. 15

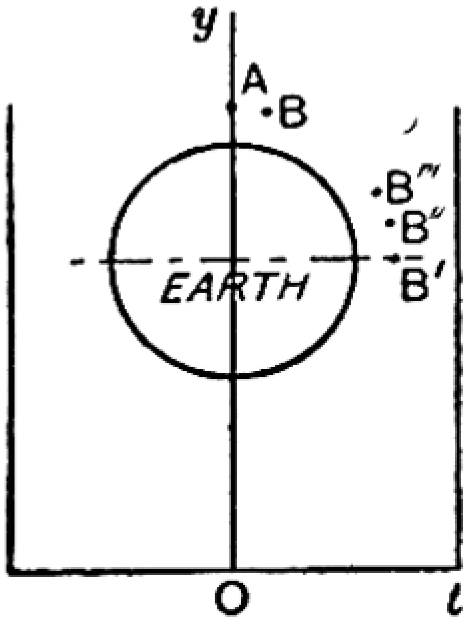

FIG. 16

The particle *A* has the space-time path *A* (Fig. 15) in this stationary lift frame. *B′* in falling describes a space path parallel to the *t* axis, and as time goes on it never gets any lower in the lift frame *yOt*, hence the space-time path of *B′* is a straight line such as *B′* in Fig. 15. Such particles as *B″*, *B‴* would have more and more curved parts, starting, however, always below curve *A* (Fig. 15), but particle *B* would be inappreciably lower in the frame *yOt* than *A*. The curve of *A* differs inappreciably from that of *B* but in examining the principle underlying the difference we deliberately exaggerate the magnitudes.

In the frame of the lift, when this is falling freely, the distance from the particle *A* to the line *Ot* (the lift floor) is constant because the particle and the frame fall freely together. Hence curve *A* (Fig. 15) becomes a horizontal straight line (*YA* in Fig. 17) parallel to *Ot*. Particle *B′* (Fig. 16), however, has no movement downwards in the frame, for it moves from right to left in Fig. 16, and so its distance from the floor (its *y* ordinate in Fig. 17) gets greater as time goes on, and we have for *B′* the path *B′* of Fig. 17.

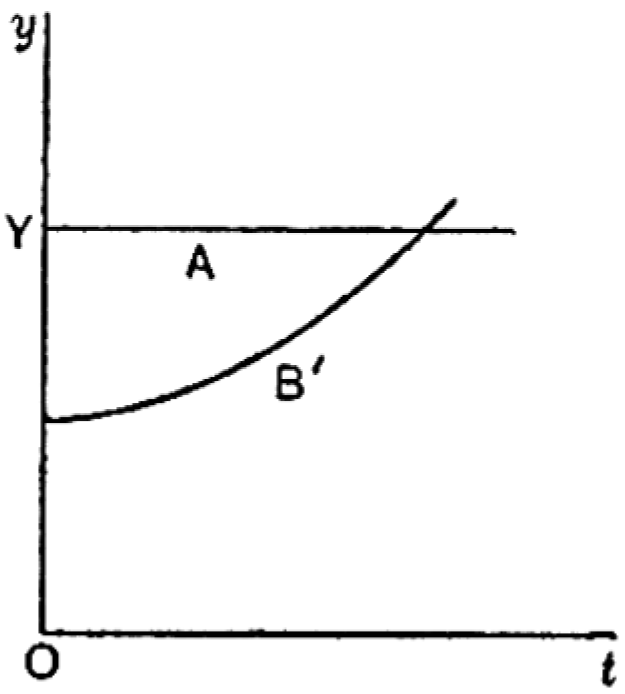

FIG. 17

Thus the use of our accelerated frame, though effective in removing the force on *A* (for it brings *A* to rest), is not effective in the removal of the force on *B′*, for it gives to this particle a bent curve in space-time (that is to say in *y*-space and time). It should be noticed that the curve *B′* in Fig. 15, though straight in *y*-space and time, is curved in *x*-space and time, but as *x*-space is measured perpendicular to the plane *yOt*, Fig. 15 fails to disclose the *x*-*t* curvature, whose presence indicates the action of gravity on particle *B′*.

Similarly it may be shown that particles *B″*, *B‴* and *B* are not brought to rest in the falling frame but have bent paths. The y-t curve for B is, of course, exceedingly little bent.

Such reasoning leads us to conclude that no possible motion imparted to the lift can remove the force of gravity from both A and B at once.

The Relativist's dilemma

Thus the Relativist is in a dilemma. He contends that "forces" are unreal. He has defined a "real" thing as that which is perceived unambiguously by all observers, and it follows from this definition that an unreal thing must be one which some specifiable observer shall be unable to perceive.

Any "force" acting on one single particle is unperceivable by that observer whose frame of reference moves so that the space-time path of the particle is a straight line in that frame of reference. Hence it is possible always to specify an observer who shall fail to perceive gravitational force at any one given point in space, but such an observer will become only the more aware of the existence of gravitational force at all the other points in the space containing the gravitating mass (e.g. the earth), for the paths of all other particles besides the one whose path has been "straightened" will appear even more bent than before. We need put attention upon no more than the two particles A and B. An observer can be specified to whom the space-time paths of both these particles appear bent—this observer says both A and B are acted on by gravity. Another observer can be specified to whom the path of A (or alternatively of B) appears straight, but to such an observer the space-time path of the other particle appears only the more bent.

The Relativist *must* find some way of making an observer see both paths as "straight." He has succeeded by an ingenious method suggested by analogy.

He specifies the observer to whom, say, A's path is "straight"—B's being bent. He asks how he can make this observer think B's path is straight also. But, is it not rather that the path must become "truly" straight? No, for straightness is a relative character, "truly" straight is a misuse of words;

all that is required is that, somehow, *B*'s path shall become "straight" relatively to an observer's frame of reference in which *A*'s path already is "straight."

The Relativist asks, what is an observer's definition of a "straight line," and finds that all individual observers as well as the universal observer agree in defining a straight line as the shortest distance between the two points called its ends. To make the bent space-time path of *B* appear straight it is necessary so to define the space-time world that the bent line is the "shortest" path between its own extremities. The distinguishing property of a "shortest" path is that it is the only unambiguously definable path. Between two points one foot apart there is an infinite choice of paths of any length greater than one foot, but there is only one path one foot long. We shall see that in space-time the path which is unambiguously definable and which by analogy might be called the "shortest" path from one "point" to another happens to be the space-time path of *greatest possible interval length*.

CHAPTER VI

The beetle

A SIMPLE analogy suggests the solution of the problem. A man once discovered a peculiar flat-backed beetle which had a remarkable instinct for finding the shortest path from place to place. This beetle seemed to be unaffected by gravity, crawling as freely up a vertical wall as across ceiling or floor. By no sort of ruse except one could he be made to follow a bent line, the ruse being to provide on the wall a hummock which should accommodate the bend in the line in such a way that this bent line formed the shortest path across the hummock. The beetle would then follow the bent line. If a series of great circles—arranged like the lines of longitude on a globe—were drawn upon an orange and the orange were then cut in halves along the plane of a great circle, or better, not quite in halves but by a section parallel to but displaced from the plane of a great circle, and the smaller "half" were then stuck on the wall so as to form a smooth hummock in the beetle's path, it was found, invariably, that the beetle traversed the hummock by way of one of the great circle paths. Gravitation had nothing to do with the matter for the beetle's course was the same whether the hummock was on wall or ceiling or floor. Presumably the great circles, though by us called bent appeared straight to the beetle. In fact the beetle seemed to be a creature whose world was confined to the 2-dimensional surfaces on which it crawled. It had no conception of the third dimension. We, perceiving the third dimension, knew that a tunnel through the hummock would be shorter than the surface-path, but to the beetle the idea of a tunnel was inconceivable; the curved path over the hummock, being the shortest surface-path, satisfied completely the beetle's idea of a "straight" path.

This beetle is just the sort of observer, in its 2-dimensional world, which the Relativist wishes to find for his 4-dimensional space-time world. Bent lines appear to the beetle to be straight provided they lie properly across a hummock—a region in the 2-dimensional world which is curved into the third dimension.

By analogy bent lines in space-time may appear straight provided they occur in a region of space-time which is "curved," and when the Relativist demands that a straight line and a bent line shall both appear straight to some specifiable observer he is demanding, in effect, that there shall be a "hummock" in space-time just in the region of the bent line.

Unfortunately we cannot visualize the "hummock," for we, in 3-dimensional space, are analogous to the beetle in his 2-dimensional space, and we can no more visualize a hummock into a new dimension than he can. Hummocks of this sort cannot be visualized but they can be described mathematically, and the manner in which they are so dealt with will have to be considered.

The reader may be inclined to ask whither such nonsense can lead, and what is the need of it? He will say that if mentally inconceivable things like hummocks in space-time are what the Relativist has to offer us in place of "forces" we prefer to retain these "forces" which are real enough to us however unreal they may be to physical science.

The Relativist replies that the reader seems to have forgotten that the object of science is to find out truth. Truth is reality, force is relative and unreal, it is not even definable. Hummocks in space-time are unambiguously definable although, as yet, they are too unfamiliar to be regarded as "possible conceptions," and they are observed similarly by all observers.

The Relativist claims to be showing us how to describe the "reality" underlying physical phenomena, and having brought us to admit that the terms "bent" and "force" are essentially ambiguous and so signify nothing "real" in nature, he proposes to invent for the description of events a new system in which such ideas as shape and force find no place. He has shown that the quantity "force" has been invented for no purpose except to "explain" the presence of a bend in the space-time path of a particle, and that a method of description in which the notion of shape has no place is one by which the ambiguous idea of force is avoided. Thus in the description of a world in which there are no "forces" (and the real world of physical science must be such) "shape" cannot be a characteristic of a particle's path. Or if, in leading up step by step from the conventional Newtonian mode of description to that of Relativity, we choose to speak of the path of some particular particle as "straight," then, if ambiguity is to be avoided, we must introduce into our description of the world, factors in virtue of which it becomes legitimate to call the path of every other particle "straight" also.

Some new definition of "straight" is thus needed, for we have seen that in no conceivable rigid frame of reference can the space-time paths of two antipodal falling particles be both at once straight in the ordinary sense.

The beetle analogy points to a way out of the difficulty. Defining a "straight line" as the "shortest surface-path" from point to point every path the beetle takes in his 2-dimensional space becomes a straight line *if and only if there do really exist appropriate hummocks* in that 2-dimensional space, namely hummocks into the third dimension (to the beetle an "impossible conception").

The Relativist will now show that if his 4-dimensional world be accepted and in place of length intervals and time intervals, only his mathematically defined "space-time intervals" be used, we shall arrive at a method of description in which there is no ambiguity and in which the space-time path of every particle is "straight"—the definition now required for "straight" being one which has reference to "space-time interval." A "straight path" in the real world of 4-dimensional space-time is like the "shortest path" in Euclidean 3-dimensional space in being the only unambiguously definable path, although, appropriately to the unfamiliar nature of most of the ideas of Relativity, this uniquely definable path (sometimes still called, by analogy, the "straight" or "shortest" path) happens to be the path of *greatest possible interval length.*

Just as, in the beetle's world of 2-dimensions, hummocks into the third dimension were necessary in order to turn "bent" paths into "shortest" paths (that is to say into paths which are unambiguously definable), so, in the Relativist's world of 4-dimensions, hummocks into the 5th dimension (an "impossible conception") are necessary in order to turn the "bent" paths of particles into *unambiguously definable* paths, namely into "geodesics" or paths of greatest possible interval length, and as no description of the world can be unambiguous (or true, therefore, to "reality") except one in which paths of particles are unambiguously definable, it follows that the "real" world of physical phenomena *must* (or strictly *may*, though until some equally satisfactory alternative world has been proposed, it is practically legitimate to say *must*) be a 4-dimensional world of space-time containing, as absolute characteristics, hummocks into a 5th dimension.

Thus in order to get over the difficulty in the problem of the antipodal particles the Relativist finds it necessary to invent a new world. This proves, however, to be a good and unambiguous world more worthy to be called "real" than any hitherto invented.

We may imagine our flat-land beetle possessed of an amount of intelligence appropriate to his shape—which was almost circular*—and able to discuss with us his behaviour towards hummocks. He would hardly repress his own conviction that hummocks were foolish nonsense and that our belief in them was due to our angularity. The beetle knew, because something at the south pole of his circular mind told him, that the idea of a third dimension in space was nonsense. He would admit fallibility but in one matter at any rate he was infallible; in his power of recognizing a straight line when he met one—and especially when he traversed one. Nature had given him a sure instinct for straight paths, and this beetle recognized immediately, even when not paying close attention, the occurrence of a bend in a path:—bends gave him a headache. As to that yellow patch with the black lines upon it, any beetle of culture would recognize at once that the paths were straight—there could be no point except that introduced by our own angularity, in calling the patch a hummock.

---

* The reader who is wise will be familiar with Dr Abbott's *Flatland*, published by Seeley & Co., London in 1884, but now out of print.

Hummocks in space

At last we hit upon a method of proving to the beetle that the hummock, though inconceivable, was "real." We asked the beetle to find for us the ratio of the circumference to the diameter of that yellow patch,—assuring him it would prove to be less than π. At first he was indignant, telling us that any fool knew the ratio must be π, and when actual measurement proved him wrong he became angry, saying the headache it had given him to measure up the circle had caused a slip in his long division.

With our knowledge of the third dimension of space it is easy to see how stupid this beetle was: the Relativist says we are similarly stupid in regard to space-time.

The reader may ask, why then does not the Relativist give us a proof of the existence of hummocks in space-time, analogous to the π proof given by us to the beetle? We at any rate can do a little mensuration and work out a sum in long division.

Non-Euclidean space

The Relativist gladly responds. He asks us to suspend a thin circle of wire by three strings from the end of a massive iron rod (a gravitating mass), as indicated in Fig. 18, making the plane of the circle contain that of the right-section which forms the flat end of the rod. We are to mark the precise centre of the circle by means of a cross on the end of the rod and then measure with precision the ratio of the circumference to the diameter of the wire circle.

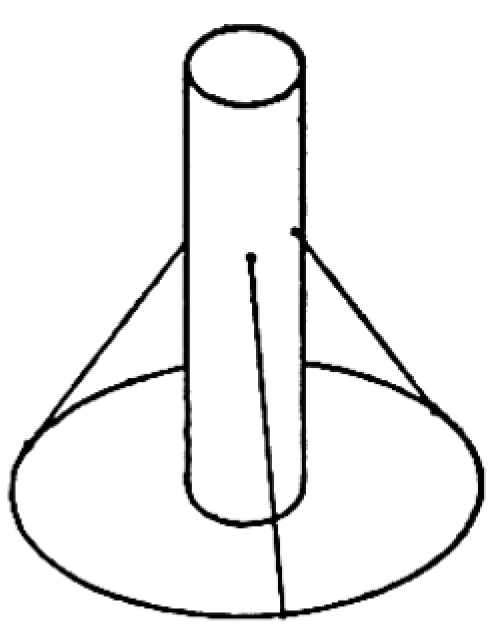
FIG. 18

The Relativist assures us that the result will be a little less than π, the defect beginning to show itself at about the 500th place of decimals!

We protest that such accuracy of measurement is impracticable, but we welcome the challenge, for we feel that, if only the Relativist can be kept down to "common-sense," we can convince him of his error without having to resort to practical measurement at all.

Our argument is simple:

We. Is the ring of wire truly circular?

Relativist. Measurement will show that all points on the ring are equi-distant from the centre point marked on the flat end of the rod.

We. Do all points of the ring lie in one plane?

Rel. If you were to remove the ring and lay it upon a Euclidean plane it would fit precisely everywhere.

We. Suppose the ring leaves its trace upon that plane, will the ratio of the circumference of that circular trace to its diameter be π?

Rel. Yes.

We. If we now replace the ring on its supporting threads and then bring our plane surface with its trace into contact with the ring will not the trace still coincide with the ring?

Rel. It will.

We. You have admitted that measurement of the trace upon the plane surface gave the ratio π. You now admit that the ring and its trace coincide. Coincidence means, in geometry, indistinguishability, hence the ring and the trace with which it coincides must have identical geometrical properties, and as the trace gave π so also must the ring. Q.E.D.

Rel. Your argument is at fault because you have assumed that when the plane was brought up against the ring and its rod it remained a Euclidean plane. As a matter of fact the gravitational field due to the massive rod produces a hummock in the space-time "around" it and when matter occupies space-time which is hummocked the matter suffers corresponding distortion.

In measuring the diameter of the ring, suspended from its threads, you would have used a material measuring rod—that is to say a rod which occupies space-time—and finding itself in a distorted region the rod would have suffered a corresponding distortion, involving an actual contraction in length when it was laid radially, that is to say along the direction of the gravitational lines of force due to the massive rod. (Here the Relativist insists on adding, parenthetically, that, strictly, what "really" happens cannot be expressed as change of length because "length" is not a "real" quantity, but to say the rod contracts is the best attempt he can make to express, in the inadequate and indeed misleading language of Newtonian mechanics, events occurring in the "real" world of physical science. Only the symbolic mathematical language of space-time intervals is adequate for describing reality.)

Having contracted, the rod would have gone too many times into that radius and so would have given a measured diameter greater in relation to the circumference of the ring than is the ratio ($1/\pi$) for Euclidean space.

Had you brought up your Euclidean plane this, acted on by the same gravitational field, would have become non-Euclidean and if you had been able then to measure the ratio of the circumference of the trace to its diameter this ratio would have proved a little less than π.

We. It is true that we had in mind a Euclidean plane made of some "rigid" substance, but this is only a detail of practical convenience. We might have measured the ring's dimensions both when it was suspended from the rod and when it was right away in empty space, by some optical method, our telescopes being then far away from both the ring and the rod.

Rel. The suggestion is a good one, and had you made it, say, in 1916, I should have replied that, although I had no experimental proof, so sure did I feel of the applicability of Relativity to every possible practical case that I would predict, confidently, the bending of rays of light in a gravitational field. The field due to the massive rod would surely bend the rays used in making the measurement, and the ratio would be found precisely that which the use of "rigid" measuring rods had already given.

To-day, however, we possess the evidence obtained by optical measurements made during the solar eclipse of May 29, 1919, which indicates that light actually is bent when traversing a gravitational field, and that the amount of the bending is precisely that required to make optical methods of measurement agree with those which Relativity says would be obtained by material measuring rods.

We. It seems to us that your principle of Relativity confuses space with things. It is true that things occupy space and that things may suffer distortion, but it seems absurd to talk of space as though it could be distorted.

Rel. To some extent your criticism is warranted. Space and time are not realities, they are modes of ordering events, they have been invented by the mind, and each individual observer has his own particular notion of what they mean. The universal observer—whose standpoint the Relativist tries to take—recognizes only a combination of space and time, and even this, strictly, is a mental invention, a mode of ordering suited to the observations which the universal observer makes of world-lines. The idea of Relativity is that there exist real factors to be related and real relations for these factors to possess. Space-time is conveniently spoken of as one of the realities of Relativity but strictly it is only the imagined "stuff" of which "intervals" are made. Space-time intervals are the only unambiguously measurable elements of the physical world and so they are the only "realities." By analogy with the common-sense notion that every real physical thing must be made of some stuff, we demand the idea of some "stuff" for intervals.

Your difficulty can be answered, however, without going so deeply into the matter.

That which is real can be measured—for the Relativity definition of reality demands that it shall be observable, and every observation consists of measurement (even recognition being a sort of measurement)—but no one ever measured space. The space separating two nails in a board means the extent of the board lying between them. The idea of space is a mental abstraction from that of things.

The Relativist's problem summarized

The clue by which the Relativist discovers the means of turning "common-sense" into "truth" is "force." In regard to "force" he finds that,

1st. Force is relative and not real, hence,

2nd. It ought to be possible to specify some observer in every system who can observe no forces in that system.

3rd. An ordinary observer defines a free particle as unacted on by force when he observes that it either is at rest or has a "straight" path in his own space-time frame.

4th. In a gravitational system (e.g. the earth's field) in which many particles are falling freely, it is easy to specify an observer in whose frame of reference any one particular particle will be at rest or moving in a straight line in space-time, but it is impossible to specify an observer in whose frame *all* freely falling particles have this property; hence,

5th. In accordance with statements 1st and 2nd it is necessary to remodel the old world of science (the Newtonian world) in such a way that the space-time which accompanies (or perhaps characterizes) gravitating matter shall so affect space-time "paths" as to "straighten" all of them at once.

6th. By analogy with a similar sort of problem laid in a world of 2-dimensional space the suggestion is a promising one that the presence of matter produces "hummocks" in space-time, the word "hummock" signifying a distortion of 4-dimensional space-time into a 5th dimension.

Following the suggestion of the 6th statement it is found necessary to remodel the idea of straight so as to fit it for use in the 4-dimensional world in which the elements of measurement are "space-time intervals," and just as in 3-dimensional space a straight path, being the shortest, is the only unambiguously definable path from one point to another, so, in our remodelled world, the world of 4-dimensional space-time, it is the uniquely and unambiguously definable "intervals" which play the part of the paths called, in the old world, "straight." In this unfamiliar remodelled world the uniquely unambiguously definable intervals, the "straight" paths, happen to be the paths of *maximum* interval length.

The position thus reached may strike the reader as eminently unsatisfactory. The Relativist has had to introduce a 5th dimension, which is inconceivable, in order to avoid "force" which is indefinable.

But why should we object to employing the inconceivable, provided it works?

If the Relativist can define a "hummock" and can show that it is unambiguously observed by all observers, he will have substantiated its "reality," and, conceivable or not, we must accept it. Unfortunately, the only language in which the inconceivable can be dealt with—except the language of mathematics, and this demands an amount of concentration and of technical skill, which only specialists possess—is the language of analogy, and no analogy can be perfect (a truism which becomes obvious when we consider that an analogy which is of unlimited range and holds in every possible detail becomes an identity, the apparent differences being only the formal and arbitrary ones of nomenclature).

We return to our beetle. His paths over hummocks were always straight, in the sense that they were the shortest, the uniquely definable paths, in the beetle's 2-dimensional world of space. We, with our knowledge of the 3rd dimension, realized that they were not "straight." We agreed, therefore, to call them "geodesics."

The beetle was proud of his geodesic instinct and proved that he could follow geodesics even with his eyes shut. There was something about the "feel" of a hummock which told him in what direction to make the next step forward. To him the geodesic path across the orange was not the complete thing which appeared to our broad vision, it was merely a *continuous succession of little steps*. The continuity was essential since it was solely by the "feel" of one step that the beetle knew how to direct his next step; had he progressed by jumps, or dared to take even one single jump, he would have missed the "feel" of the interval thus bridged and so would have chanced his orientation.

As we wanted to check very exactly his claim to geodesic instinct we had to follow and record the beetle's steps with great care. The method used was to draw a network of lines upon the surface of a hummock, the lines (which we may call "threads") bearing distinctive numbers by aid of which every mesh in the network could be identified. The beetle's path was then described by a record of the meshes traversed. Obviously, provided the meshes were small, the shapes and spacings of the "threads" might be chosen at random, we could use any mesh-system we pleased.

Having recorded the beetle's path by aid of our arbitrary mesh-system we afterwards traced the path in pencil and tested it to see whether or not it was a true "shortest" path or geodesic. For very smooth hummocks which were convex all over, this test could be made by stretching a well lubricated thread from the beginning to the end of the path, but in other cases the test had to be made mathematically.

The mathematical problem involved was not simple. Our record of the path consisted of a series of numbers indicating, in proper order, definite meshes of the network which we had drawn upon the hummock, but as the "threads" of this network had been laid on at random and contained erratic bends, the relation between the meshes and the hummock was in no sense discernible from the numbering, but each mesh had to be found by actual search over the surface of the hummock. The length of path between two marked points on the hummock bore no definable relation to the number of meshes traversed by the path, for the meshes, though deliberately made small, had all sorts of odd sizes and shapes.

Even in a plane surface it would be tedious to define a rectilinear path in terms of the numbered meshes. Thus two factors had to be taken into account in finding a geodesic on a hummock, namely, the nature of the mesh-system and the curvature, from place to place, of the hummock.

What a wonderful mathematical instinct the beetle thus appeared to possess! It made no difference what mesh-system was marked out—nor indeed whether any was marked out at all—he knew by the "feel" of each infinitesimal area of surface everything necessary for the definition of geodesics. He was a tiny fellow, covering an area so small that it might be indicated appropriately by $(ds)^2$, where ds is an infinitesimal interval of path. It was surely by observations made upon such little areas as ds^2 that the beetle arrived at the definition of each step in a geodesic and was able also to string together in proper order the series of intervals ds of which a geodesic was composed.

Mesh-systems

This thought suggested that we also, by careful measurement of and around each ds over an area ds^2, ought to be able to determine the presence and shape of any hummock.

As an introduction to the properties of space intervals upon curved surfaces it was customary first to consider those of intervals upon plane surfaces. In order to be able to specify any particular small region in a plane surface, it is usual to lay on the surface a mesh-system. One particular sort of mesh-system is in common use; it would be used, for instance, in specifying some particular letter on this page. The letter "*g*" just written in the word "page" can be located from the information that it is in line 24 and column 45, it being understood that the "columns" are vertical channels each one letter wide.

The mesh-system used in such a case consists of a number of equally spaced "horizontal" lines crossing a number of equally spaced "vertical" lines. Vertical distances, measured in spaces, are usually called "y" and horizontal distances "x."

In such a mesh-system on a plane surface the square of the straight line "s" (Fig. 19) joining any two points (x_1, y_1) and (x_2, y_2) is given by $s^2 = (y_2 - y_1)^2 + (x_2 - x_1)^2$ however great s may be.

Other mesh-systems are used, as, for instance, a system of straight lines y and x which are inclined at some angle other than 90°. The equation for s^2 is then

$$s^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + 2\,(x_2 - x_1)\,(y_2 - y_1)\cos\theta$$

where θ is the angle yOx (Fig. 20).

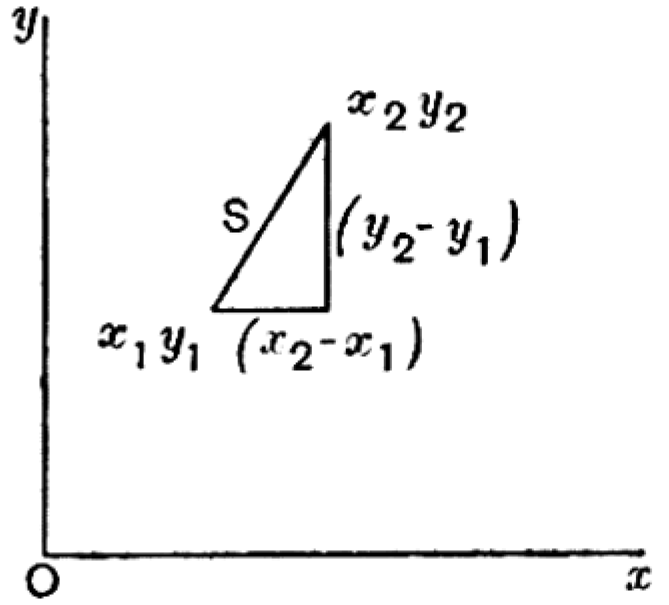

FIG. 19

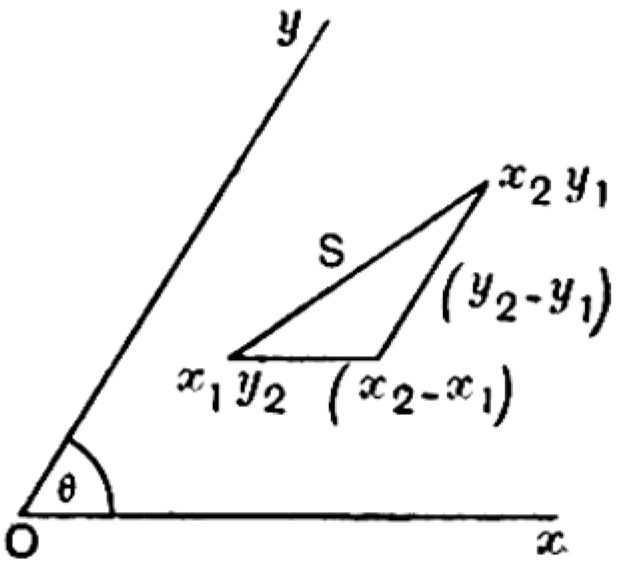

FIG. 20

Another well-known system consists of a series of equidistant concentric circles, having the "origin" (O) as centre, intersected by equally spaced radii (Figs. 21 and 28). In this system the position of a point P on the surface is specified by stating its distance "r" from O, and the angle "θ" separating its "r" from the fixed line OR.

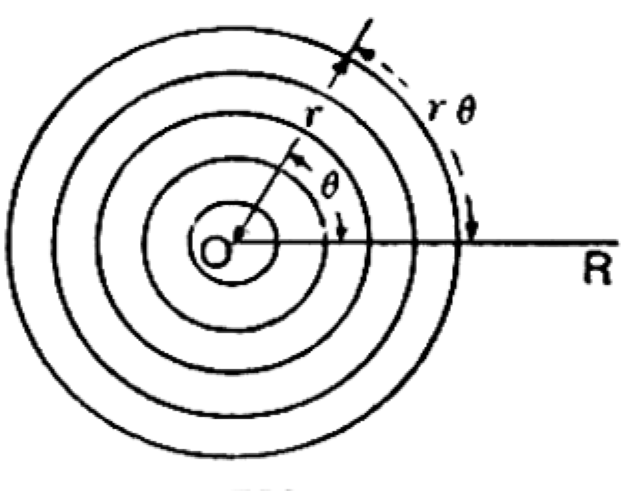

FIG. 21

Mesh-systems of this sort are much used in connection with temperature or voltage or pressure recorders, in such a way that θ, the angle turned through by the chart, is proportional to time.

This sort of chart (the co-ordinates r and θ of which are usually called "polar" co-ordinates because they specify position in relation to distance from the pole O and angle turned about it) introduces a difficulty common to all mesh-systems using bent lines. The formula

$$s^2 = (y_2 - y_1)^2 + (x_2 - x_1)^2$$

expresses Proposition 47 of Euclid, Book I, and shows the relation between the sides and the hypotenuse of a right-angled rectilinear triangle.

The $\cos\theta$ formula also is applicable only to triangles which are rectilinear. We cannot apply these formulae without error to triangles having bent sides.

It is an interesting fact, however, that the error introduced by using the formulae for triangles with bent sides is much less for small triangles than it is for large ones. This may easily be proved: we consider the triangles *OBC* and *OAD* (Fig. 22) and have to show that the percentage error in writing

$$(ODC)^2 = OB^2 + BC^2,$$

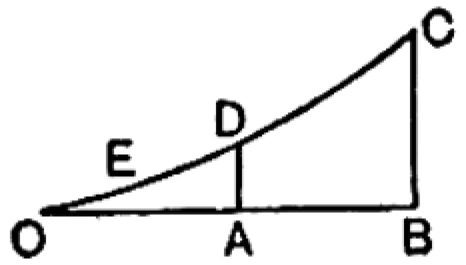

FIG. 22

that is $(ODC)^2 = (OC)^2$ for the larger triangle, is greater than the percentage error in writing

$$(OED)^2 = OA^2 + AD^2,$$

that is $(OED)^2 = OD^2$.

In other words we have to show that if the length of the curve *ODC* exceeds that of the straight line *OC* by, say, 10 %, then the curve *OED* will exceed *OD* by less than 10 %.

It will suffice to consider the special case in which *ODC* is a circular arc and *D* is its mid-point. Then since the lines $(OD + DC) > OC$ the percentage excess of the arc *OEDC* over the lines $(OD + DC)$ must be less than the excess of the same arc over the straight line *OC*; hence the excess of $\frac{1}{2}(OEDC)$ over $\frac{1}{2}(OD + DC)$, that is of *OED* over *OD*, is also less than the excess of *OEDC* over *OC*, and as *OEDC* exceeds *OC* by 10 % then *OED* must exceed *OD* by less than 10 %.—(Q.E.D.)

Arguing in a similar way it may be shown that when a triangle with a curved hypotenuse is exceedingly small (by which is meant that the ratio of the length of the hypotenuse of the triangle to its radius of curvature is negligible compared with unity), the error in writing $s^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2$ is less than can be measured. And since the equation in this form is the equation of Euclid I. 47 the same fact may be expressed by saying that exceedingly small portions of continuous curved lines have the properties of Euclidean straight lines.

The principle may be extended to surfaces and we may say that geometrical figures drawn upon curved surfaces have Euclidean properties if the figures are very small.

In order to indicate, in the case of triangles with curved sides, that very small triangles are being considered, it is customary to write ds instead of s and dx and dy instead of $(x_2 - x_1)$ and $(y_2 - y_1)$ respectively.

Applying the ds^2 equation to a small length interval in a plane surface laid out in polar meshes, we may write (see Fig. 23):

$$(ds)^2 = (dr)^2 + (rd\theta)^2$$

or, $ds^2 = dr^2 + r^2d\theta^2$, or writing x for r and y for θ:

$$ds^2 = dx^2 + x^2dy^2.$$

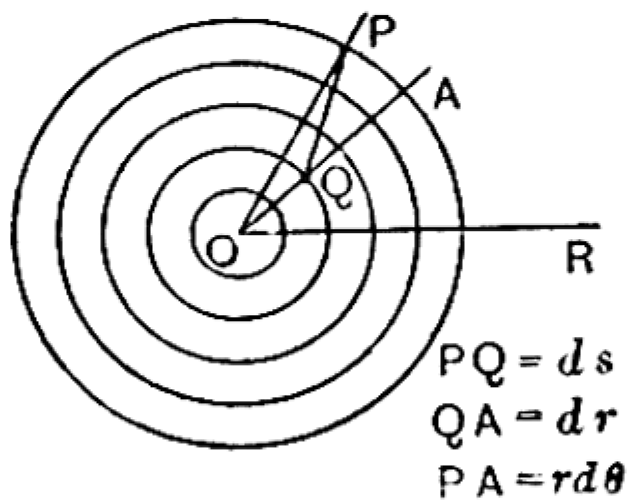

FIG. 23

We have dealt with three of the common sorts of mesh-system used on plane surfaces, and have given the formula for ds^2 in each case, namely:

For rectangular meshes,

$$ds^2 = dx^2 + dy^2. \qquad \text{.........I}$$

For oblique meshes,

$$ds^2 = dx^2 + dy^2 + 2dx\,dy\cos\theta. \qquad \text{.........II}$$

For polar meshes,

$$ds^2 = dx^2 + x^2dy^2. \qquad \text{.........III}$$

Mathematicians have investigated the problem of mesh-systems in general and have discovered that the ds^2 equation for any mesh-system whatever—consisting of a warp and a woof—without restriction as to the shapes of the meshes or the spacings of the y and x threads, may be put into the form:

$$ds^2 = g_{11}dx^2 + 2g_{12}dx\,dy + g_{22}dy^2, \qquad \text{.........IV}$$

or employing, as is customary in the general case, x_1 instead of x and x_2 instead of y:

$$ds^2 = g_{11}dx_1^2 + 2g_{12}dx_1dx_2 + g_{22}dx_2^2. \qquad \ldots\text{IV}$$

In this equation the g's, which are multipliers of the dx products, may be either fixed numerals or else "functions" of x or y (that is to say of x_1 or x_2). A function of x is a quantity whose value depends upon the value of x. Equation IV gives the geometry of a small triangle located *anywhere* in the mesh-system, the location being determined by the values arbitrarily chosen for x_1 and x_2; thus to say that in some mesh-systems the g's are functions of the x's is to say that in some systems (including, obviously, those in which the threads are laid down erratically) the geometry varies from spot to spot of the system and that for each locality—that is to say for each chosen value of the x's—values peculiar to that locality have to be assigned to the g's.

Introduction to the "g's"

In cases like I and II, in which all the meshes are alike, the equation for ds^2 which applies in any one locality of the meshed plane surface applies in every locality; that is to say in such cases the g's are constants. When, however, as in III and in cases of random mesh-systems, the shapes of meshes differ in different regions (localities), the ds^2 equation must, somehow, depend upon the region. In these cases some or all of the g's must be functions of the x's.

The values of the g's of equation IV which turn this general equation into the special forms I, II and III respectively, are seen by inspection to be the following:

	g_{11}	g_{12}	g_{22}
In equation I	1	0	1
,, II	1	$2\cos\theta$	1
,, III	1	0	x_1^2

For ease of inspection we write out again the general equation and under it equations I, II and III:

$$\begin{array}{rlcl} & ds^2 = g_{11}dx_1{}^2 + g_{12}dx_1dx_2 & & + g_{22}dx_2{}^2. \\ \text{I} \ldots & ds^2 = \quad dx_1{}^2 & & + \quad dx_2{}^2. \\ \text{II} \ldots & ds^2 = \quad dx_1{}^2 + 2dx_1dx_2\cos\theta & & + \quad dx_2{}^2. \\ \text{III} \ldots & ds^2 = \quad dx_1{}^2 & & + x_1{}^2dx_2{}^2. \end{array}$$

One special point of difficulty in connection with these ds^2 equations shows itself in one's inclination to suppose that since very small triangles, even though their sides be bent, may be treated as rectilinear, the properties of the mesh-system cannot be contained in the geometry of these small triangles—at any rate such a tiny triangle cannot know whether it lies in a mesh-system having straight threads or bent ones.

This difficulty disappears as soon as we consider exactly what practical procedure is followed when we set about to examine a mesh-system whose properties are unknown.

We are dealing at present only with mesh-systems laid out upon plane surfaces. Let Fig. 24 be such a system. We assume that the mathematicians are right when they say that in any mesh-system whatever the equation for ds^2 must be of the form

$$ds^2 = g_{11}dx_1{}^2 + g_{12}dx_1dx_2 + g_{22}dx_2{}^2.$$

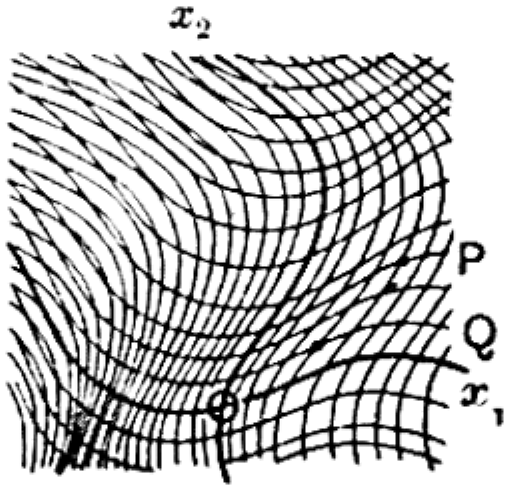

FIG. 24

How are we to find the values of the g's for Fig. 24?

We choose at random any pair of threads Ox_1 and Ox_2 crossing at a point O which shall serve as "origin" of co-ordinates, and use these as axes of reference. We then number the threads of the warp (by which are meant the x_1 threads) 0, 1, 2, 3 ... or in any arbitrary manner provided that the threads which lie in succession be given numbers whose magnitudes are in succession. Then the woof threads must be numbered in corresponding fashion.

Since the thing which has to be described in terms of the properties of these threads is a *continuous* surface, the threads must form a *continuous* envelope for that surface. The threads actually drawn and numbered are only sample threads picked out for reference; the system to which these threads belong must be a continuous system having in its warp and also in its woof an infinite number of consecutively contiguous threads. Each of these infinitely numerous threads must bear a distinctive number and the numbering must be such that contiguous threads bear contiguous numbers, that is to say threads which are in very close succession must bear numbers which also are in similar close succession. The only practicable approach to the representation of this is to number with successive numbers (conveniently, whole numbers) successive sample threads, it being understood that between every pair of successive sample threads there lie an infinite number of other threads forming a continuous non-overlapping series, and between the whole numbers assigned to two successive sample threads is the infinite continuous series of fractional numbers, one of which is thought of as assigned to each of the imagined threads.

The threads of reference Ox_1 and Ox_2 having been marked and the x_1 and x_2 threads numbered any point such as P (Fig. 24) may then be specified unambiguously by giving the numbers of the warp and of the woof threads at whose intersection it lies. Should P lie off one of the marked sample threads the number of the imagined thread on which it lies has to be estimated, and the assumption is that so closely have the sample threads been laid or so uniformly have they been spaced that the required estimate can be made with sufficient practical precision.

It is usual, in specifying a point, to give first the number on the end of the "vertical" thread—that is to say the value of x (or x_1) for the point—and then the number on the end of the "horizontal" thread—that is to say the y (or x_2) of the point. Thus the point P in Fig. 24 is (x_1, x_2) where

$x_1 = 5$ and $x_2 = 3$, while Q is (x_1', y_1') where $x_1' = 4$ and $x_2' = 1$.

For clearness we have taken P and Q far apart in Fig. 24, but in order to determine the "g's" we must take two points *very close together*. What *very close together* means depends (in the case of plane surfaces) upon the curvature we have chosen to give to the threads of our mesh-system. If the curvature is great—that is to say if the bends in the threads are sharp—the points must be nearer together than when the curvature is small. This means that the distance between P and Q should be practically negligible in comparison with the radius of a circle which could be made to fit the curved mesh-lines in that region. Whether or not a quantity is practically negligible depends upon the precision of the instruments we are using for the measurements. In measurements made with a draper's yard-stick differences of, say, 1/64th of an inch are practically negligible because the stick itself may well have errors of marking greater than 1/64th of an inch.

If we are recording an experiment in which the instruments used have no greater accuracy than, say, 1 %, then it will probably be of no advantage to take the distance PQ smaller than, say, 1 % of the radius of curvature of the mesh-lines in that region.

That the rule suggested, namely, to take the distance PQ equal to that percentage of the radius of curvature which gives the percentage accuracy of our instruments, is surely on the safe side, may be seen by considering Fig. 25 in which the points P and Q are shown upon one of the mesh-lines of the system (actually a warp thread because in this region the warp threads have greater curvature than those of the woof).

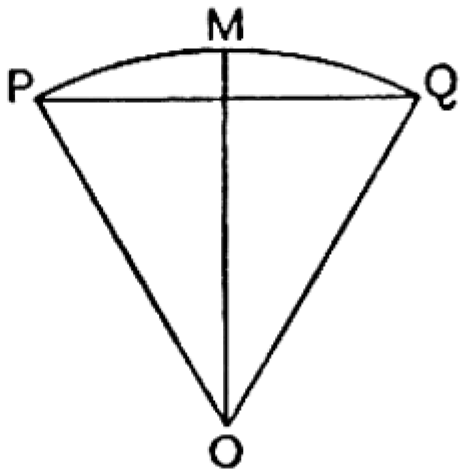

FIG. 25

The error in speaking of the straight line PQ as (ds) is given by the difference between PQ and the arc PMQ. If O is the centre of the circle which practically fits the arc PMQ then it is clear by inspection that the percentage that *PQ* is of the radius of curvature *OP* is far greater than the percentage which *PMQ* minus *PQ* is of *PMQ*.

With such an understanding of the meaning of "near together" we take then the two points *P* and *Q* near together in any region of the chart in which we wish to find the "*g*'s." We then measure the straight line *PQ* in any convenient units, say inches, and we write:

$$ds^2 = g_{11}dx_1{}^2 + g_{12}dx_1dx_2 + g_{22}dx_2{}^2.$$

Measurement gave $PQ = ds =$ say, 1·232 inch and we had for *P*, $x_1 = 5$ and $x_2 = 3$, while for *Q* we had $x_1{}' = 4$ and $x_2{}' = 1$, whence $dx_1 = (x_1 - x_1{}') = (5 - 4) = 1$ and

$$dx_2 = (x_2 - x_2{}') = (3 - 1) = 2,$$

the *x*'s being simply the arbitrarily assigned numbers borne by the mesh threads concerned.

The equation thus becomes,

$$(1{\cdot}232)^2 = g_{11}(1)^2 + g_{12}(1 \times 2) + g_{22}(2)^2.$$

As there are three unknowns we require two more such equations *involving the same g*'s, and obtained by measurements made in the same region. We choose, therefore, two other pairs of points also very close together and both pairs close to *P* and *Q*. Usually it will suffice to make *P* or *Q* common to the other two pairs of points, the first (*ds*) being *PQ*, the next *PQ′* and the third *PQ″*. Proceeding exactly as in the case of *PQ* we find two additional $(ds)^2$ equations and so solve for the three unknowns g_{11}, g_{12} and g_{22}.

The *g* suffixes 11, 12 and 22 are used because the symbols x_1 and x_2 being preferred to *x* and *y* the suffixes in these forms serve well to suggest to which combination of the *x*'s each particular *g* belongs. Thus g_{11} is the multiplier of $(x_1 \times x_1)$ or $x_1{}^2$, g_{12} the multiplier of x_1x_2 and g_{22} the multiplier of $x_2{}^2$. The advantage of x_1 and x_2 over *x* and *y* is chiefly that the system is more uniform when continua of more than two dimensions are being dealt with; for instance, in the 4-dimensional continuum of space-time the symbols x_1, x_2, x_3, x_4 are used in preference to x, y, z and t.

These then are the g's for the region in which P, Q, Q' and Q'' lie. This region is small, its area being that of a circle having P as centre and (ds) as radius, and (ds) is very small. We define the region, therefore, with good precision by saying it is the region (x_1, x_2) where x_1 is the arithmetic mean of the x_1's and x_2 that of the x_2's for P, Q, Q' and Q''.

We record the g's just found as the g's (or "potentials") of the mesh-system in question, in the region (x_1, x_2); and by continuing the process the g's for other regions may be found.

The gradient of the g's

In regions where the curvatures of the mesh-lines are different from the curvatures at (x_1, x_2) the g's will be different, and it is not surprising to find that the rate of change of the g's around a given region depends upon the rate of change of the curvatures of the mesh-lines.

A 3-dimensional chart may be plotted showing the way in which g_{11} varies with the region in any 2-dimensional mesh-system, another of g_{12} and a third for g_{22}.

Any quantity so related to other quantities that the relation can be represented unambiguously and definitely by means of a continuous "curve," is said to be a function of these other quantities. The word "curve" is here used in a broader sense than usual, and includes not only curved lines (1-dimensional curves) but also curved surfaces.

Fig. 26 is intended to indicate the way in which the connection between the various regions of a meshed plane and the values of one of the g's (say, for instance, g_{11}) might be represented by a curved surface. At each region of the meshed plane a perpendicular is erected of length equal to the value there of the g_{11}. When an unlimited number of regions have had their proper vertical lines assigned to them the tops of this forest of lines

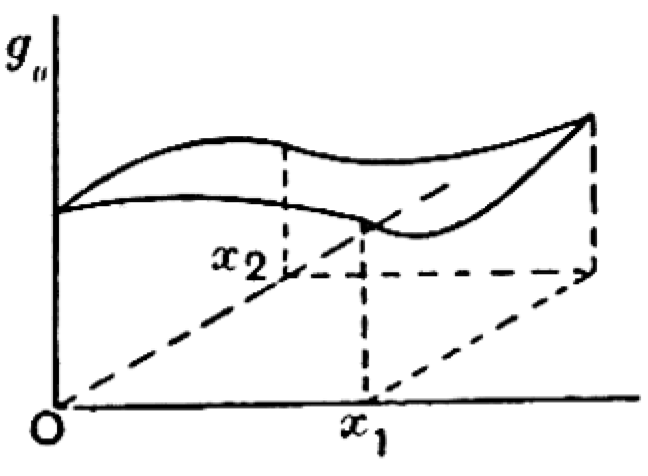

FIG. 26

will occupy a surface, which, in general, will have hills and valleys in it. For clearness the meshes in the chart x_1, O, x_2, have been omitted.

The various gradients upon this *g* surface of hills and valleys are called the gradients of the *g*'s; they depend upon the rates of change of the curvatures of the mesh-lines in the corresponding region of the mesh-system.

The *g*'s can always be represented by continuous curves of this kind (provided the lines of the mesh-system are continuous and always are smoothly curved), that is to say the *g*'s always are a function of x_1 and x_2, but when a *g* surface proves to be a plane parallel to the plane x_1Ox_2 it is usual to call *g* a constant.

We are now in a position to see why it is that the properties of triangles which are so small that their sides, though lying along curved mesh-lines, may be considered straight, do indeed disclose the nature of a mesh-system having curved mesh-lines. One alone of such little triangles does *not* disclose the nature of the system; it is necessary to examine numerous (*ds*) triangles in the immediate neighbourhood of each region. Thus the *gradients* of the *g*'s are discovered and it is the *g*'s *and their gradients* which disclose the nature of the mesh-system.

The first step has now been made towards the determination of geodesics. It seems disheartening that we have to go to so much more trouble than the beetle. He cares nothing about the subject of mesh-systems, which we have to master before we begin upon that of geodesics, and it is indeed a fact that the subject of mesh-systems is in no essential way relevant to that of geodesics.

Our object is, however, very different from the beetle's. We wish to find out and record an exact description of our world's geodesic structure: this is a much wider object than that of simply following a geodesic path—every star and planet does that and cannot help doing it.

The description has to be written in the mathematical language of mesh-systems.

We have pictured mesh-systems in a plane surface. Sometimes, in practical demonstrations before a class of students, the mesh-system employed consists of a shadow thrown by a lantern upon a screen. Usually it is the shadow of the image on a photographic plate; it might, however, be that of a piece of wire gauze. We will imagine that the network on the screen is the shadow of a sheet of wire gauze which has been shaped into a hummock, so that it is like a miniature dish cover. The image projected on the screen is a mesh-system—probably very irregular—described upon the plane surface of the screen.

We will suppose that the *g*'s for this screen image have been found, the choice of axes Ox_1 and Ox_2 as well as the numbering of the lines of the mesh-system having been reproduced by corresponding markings upon the actual gauze.

Attention is turned now to the mesh-system provided by the actual gauze with its hummock. Points P, Q, Q' and Q'' having been chosen for *ds* measurements upon the screen image, these same points will be used upon the hummocked gauze, for unless the hummock has very steep and erratic curvature into the 3rd dimension, points whose projected images were "near together" on the screen will also be "near together" on the gauze.

For convenience we suppose the projection on the screen to have been made without magnification, and first we note that distances (*ds*) from P to Q, P to Q' and P to Q'' measured now along the hummocked surface of the gauze will be greater than the projections of these which were measured on the screen. Hence the ds^2 in the gauze equations will differ from the ds^2 of the *otherwise similar* screen equations, and therefore the *g*'s for the gauze will differ from the *g*'s for the screen image.

The *sole cause* of this difference between the screen *g*'s and the gauze *g*'s is the hummock, for the very same mesh threads have been used for both and the same thread numbering.

We may next imagine that a *precisely similar* hummock is constructed of gauze having some peculiar irregular mesh—though still consisting, of course, of a warp and a woof of continuous untangled threads. The screen image gives now a set of *g*'s different from those given by the image of the first gauze, and these differences express the change in mesh-system. The actual new gauze gives a new set of gauze *g*'s, and the differences between these *g*'s and those of the new screen image express the hummock.

This process of changing the mesh may be continued indefinitely. We then have a long series of different groups of *g*'s for the screen images, and another long series of different groups of *g*'s for the actual gauzes. And these two series of groups are related in a definite manner, because the differences between their corresponding elements (screen *g*'s and gauze *g*'s) express always the same hummock.

That is to say there is some "invariant" property common to the sets of *g*'s of all the long series of gauze equations, which expresses the presence of that particular hummock, and which is absent from all the sets of *g*'s of the screen equations.

CHAPTER VII

The gradient of the gradient of the g's

We have been led to the conclusion that there must be in general some characteristic in the group of g's found for a 2-dimensional continuum (that is to say, for a surface) which expresses the presence and nature of any hummock that may be present, and that this characteristic must be discernible in the group of g's quite independently of the arbitrary choice of the mesh-system which may have been used.

It is hardly surprising that the determination of this characteristic requires advanced mathematical knowledge and skill.

Fortunately for scientific progress men have not been wanting whose love of thought for its own sake led them, long before experimental science had reached the point at which it could apply the results of their work, to develop the mathematics of mesh-systems in general, with and without hummocks. Einstein has been able to make use of the work of Christoffel, Gauss, Helmholtz, Levi-Civita, Riemann and others, without which it is probable that the development of the Theory of Relativity would have been long postponed.

It is probably impossible to explain in non-mathematical language just what property of the g's it is which discloses a hummock, but a little light may perhaps be thrown on the matter by consideration of one simple special case, namely the comparison of the screen image with the gauze system of a symmetrical hummock shaped like the top of a parrot's cage (see Fig. 27). The gauze or wire work consists of a series of wire circles lying in parallel horizontal planes and having their centres on a common vertical axis, kept apart by a series of equally spaced ribs arranged like the ribs of an open umbrella and crossing the circles always at right angles.

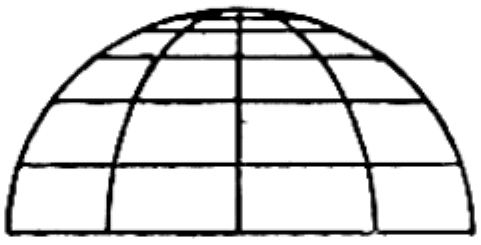

FIG. 27

We imagine the spacing of the circles to have been chosen so that a parallel beam of light sent with its rays parallel to the axis of the cage produces on the screen a system of polar co-ordinates (Fig. 28), that is to say a mesh-system like a spider's web but having the circles equally spaced along the radii while these latter are equally spaced around the circles.

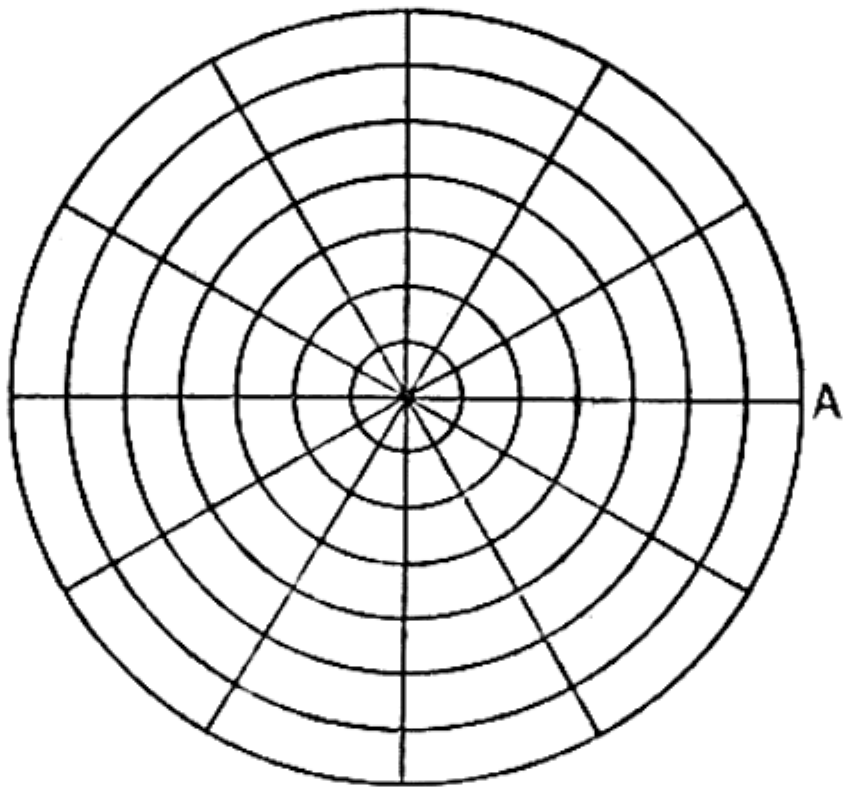

FIG. 28

In this system the x_1 lines are the circles and the x_2 lines the radii. The radii may be numbered continuously by labelling them with the values of their angular distances (say in radians) from the horizontal line OA, O being the centre of the circles:—this will be in conformity with the usual custom in regard to polar co-ordinate systems, except that instead of r and θ we are using the symbols x_1 and x_2.

We know from the geometry of Fig. 28 that

$$ds^2 = dr^2 + r^2 d\theta^2,$$

i.e.

$$ds^2 = dx_1{}^2 + x_1 dx_2{}^2,$$

and have seen already that the g's for this mesh-system on a plane surface are:

$$g_{11} = 1, \ g_{12} = 0 \text{ and } g_{22} = x_1.$$

The charts showing the values of these g's are therefore: for g_{11} a horizontal flat surface unit distance above the mesh-system (Fig. 29); and for g_{22} a conical surface as in Fig. 30, enclosing a solid angle of 90°.

It is instructive to notice the common-sense reasons why in this case the g's have the values we have found.

In plane rectangular mesh-systems big (s) triangles as well as little ones are rectilinear and right-angled and the equation for them all is

$$b^2 = a^2 + c^2,$$

which, written for a very small triangle, is

$$ds^2 = dx_1{}^2 + dx_2{}^2,$$

while in oblique rectilinear systems the equations are

$$b^2 = c^2 + a^2 - 2ac \cos B,$$

or

$$ds^2 = dx_1{}^2 + dx_2{}^2 - 2dx_1 dx_2 \cos \theta,$$

where θ is the angle B.

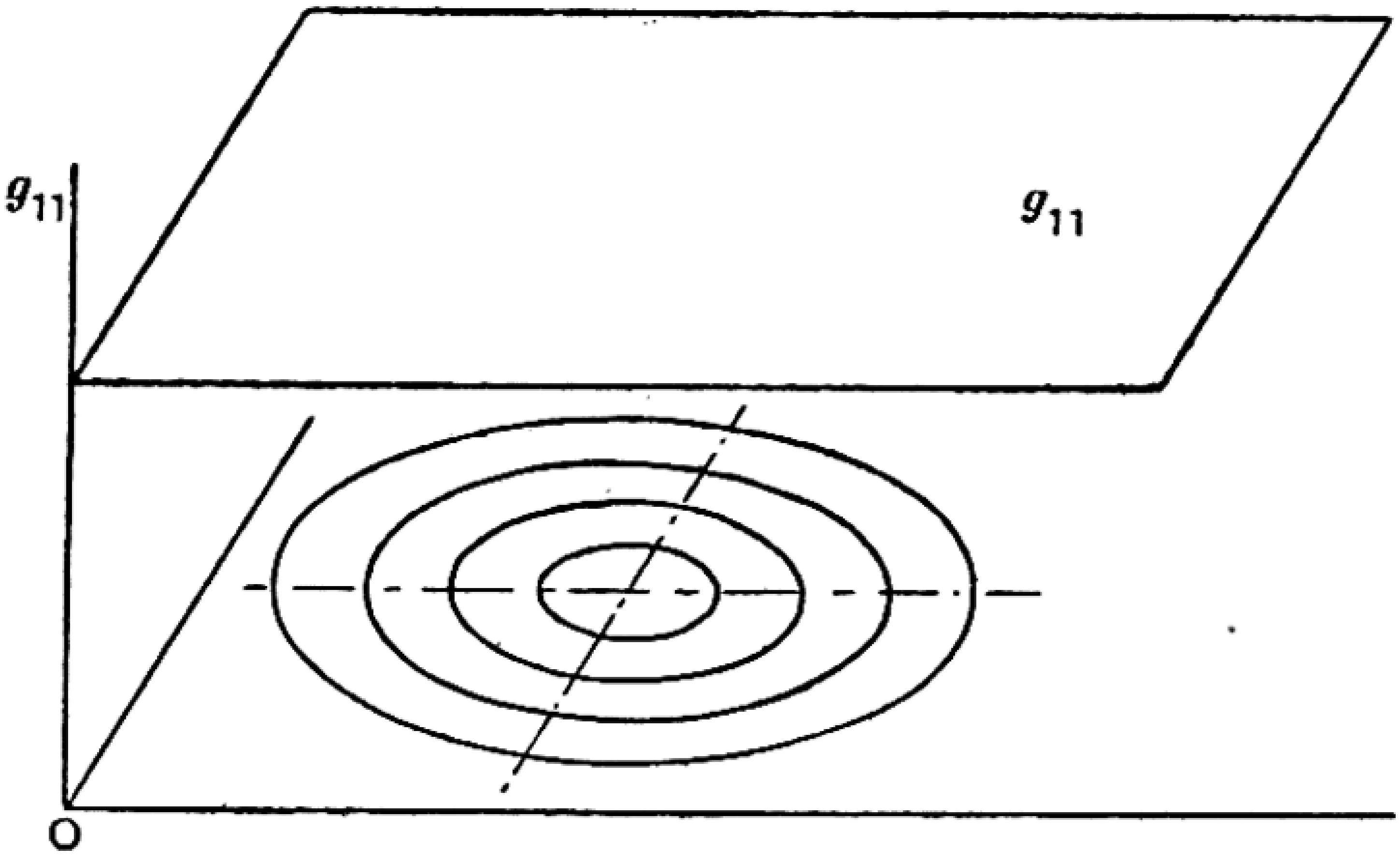

FIG. 29

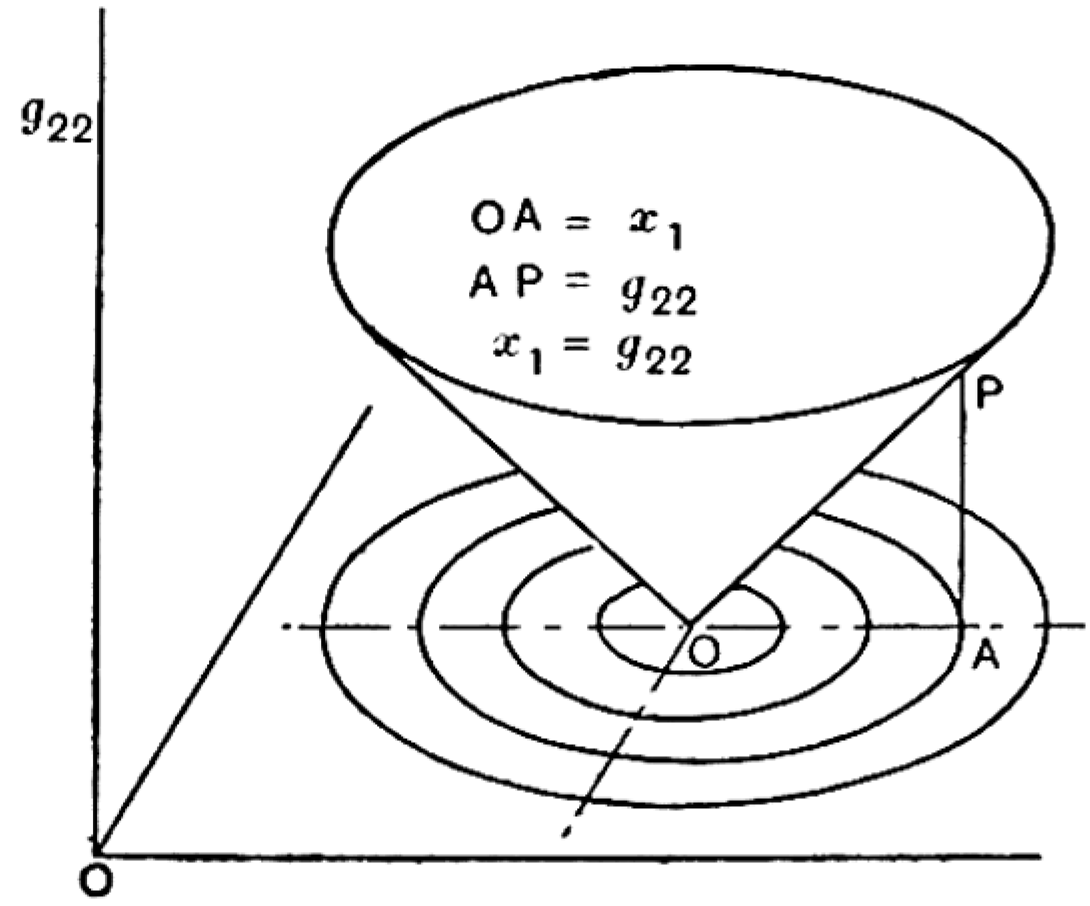

FIG. 30

In all these equations it is assumed, of course, that the hypotenuse of the triangle is measured in the same units as the sides, and it is for this reason that g_{11} and g_{22} (the multipliers of dx_1^2 and dx_2^2 respectively) are each unity.

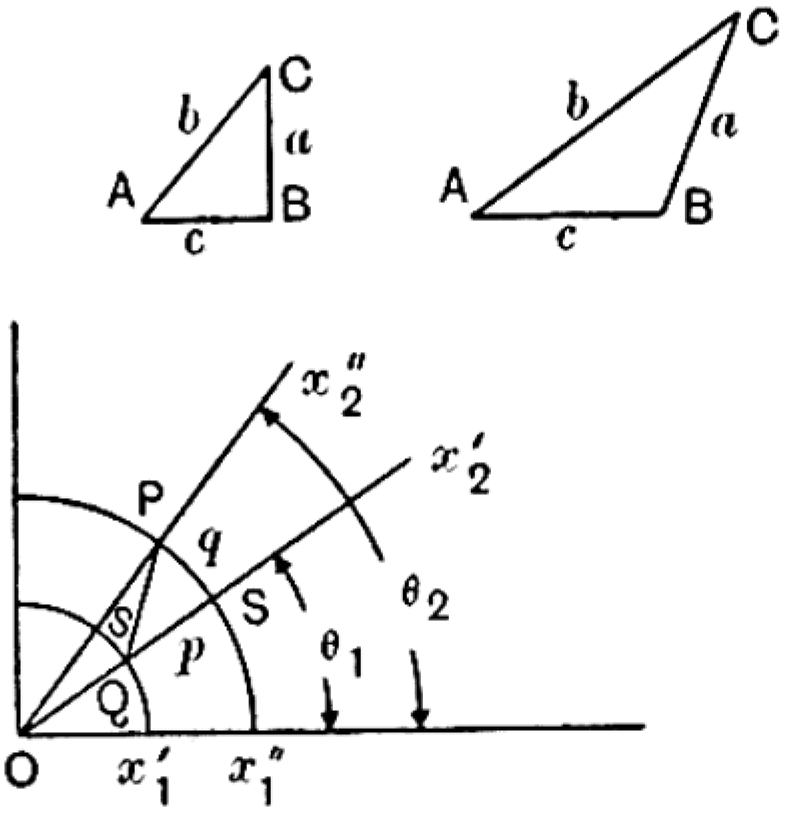

FIG. 30 A

In our polar system the small triangles may be regarded as rectilinear, and as they are rectangular we may write

$$s^2 = q^2 + p^2$$

provided s, q and p are small and provided also that they are measured in the same units.

When, however, following the method given for finding the g's, we label the radii OP and OS with numbers x_2' and x_2'' signifying the circular angles θ_1 and θ_2, and the circles through Q and P we label x_1' and x_1'' respectively, it becomes necessary to pay careful attention to the question of *units if we wish to control the numerical value of the* g*'s.*

Thus the equation $s^2 = p^2 + q^2$ becomes, simply,

$ds^2 =$ the square of the arc lying between the radii x_2'' and x_2' + the square of the radius lying between the circles x_1'' and x_1', provided ds and the arc and the radius are measured all in the same units, say, inches, or are so labelled. But, following the usual custom we have $(x_2'' - x_2') = (\theta_2 - \theta_1)$ *radians* and to bring it to inches it must be multiplied by the radius OP in inches, that is to say by x_1''. Conforming then to this condition as to the labelling of the mesh-lines we write

$$ds^2 = dx_1^2 + x_1 dx_2^2,$$

having controlled the units so as to make g_{11} unity.

g_{22} is then x_1 and g_{12} disappears because, our ds triangles being all rectangular, θ is 90° and $\cos\theta = 0$.

This consideration is important because it shows that the values of the g's for a given mesh-system depend upon the labelling arbitrarily chosen for the threads and also upon the units of length chosen for the measurement of the (ds)'s.

For the purpose of finding hummocks it is of no importance what mesh-system has been chosen, nor how the threads are labelled, nor in what units (ds) is measured, provided the rule of continuity is adhered to, by which it is ensured that for any pair of points (x_1', x_2') and (x_1'', x_2'') which are "near together" in the system the values of $(x_2'' - x_2')$ and $(x_1'' - x_1')$ shall be small, that is to say $(x_2'' - x_2')$ shall be a practically negligible fraction of either x_2'' or x_2', and similarly for $(x_1'' - x_1')$. It is well to mention this simple point because it is at variance with the foolish but common convention by which two numbers whose arithmetic difference is the same are said to be equally near together. With our understanding of the meaning of a "small difference" and of "nearness together" two such numbers as, say, 1000 and 1001 have a much "smaller difference" and are much "nearer together" than, say, 100 and 101, although their arithmetic difference is the same.

We return to our parrot cage. For simplicity we label its wires and measure our distances in such a way that the equation for the (ds^2)'s of the projected image is

$$ds^2 = dx_1{}^2 + x_1 dx_2{}^2$$

and $g_{11} = 1$, $g_{12} = 0$, $g_{22} = x_1$.

Next we find the g's for the wire cage itself.

The hypotenuses (ds) and the sides of our (ds) triangles lie now upon a hummock and for any wire triangle whose projected image gave

$$ds^2 = dx_1{}^2 + x_1 dx_2{}^2$$

we now have $$ds^2 = dx_1{}^2 + k'' x_1 dx_2{}^2,$$

where $k''x_1$ is the radius of the wire circle on which P lies, x_1 being now something greater than this radius, since it is the actual length of the curved rib of the cage on which P lies.

The equation shows that g_{11} and g_{12} will still have their screen image values of 1 and 0 respectively but that g_{22} will be $k''x_1$ instead of x_1. Fig. 31 shows a section of the cage taken through the plane containing one of its ribs, the sections of the individual wire circles being indicated by dots.

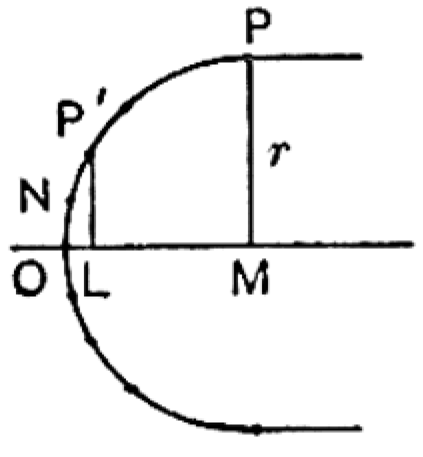

FIG. 31

For a (ds) triangle at P the value of $k''x_1$ will be $PM\ (= r)$ where x_1 means the length of rib from O to P. In such a case (Fig. 31) $k''x_1\ (= g_{22})$ is about $(x_1/2)$ and $k'' = \frac{1}{2}$, since PM is about $\frac{1}{2}OP'P$. But for a triangle nearer the "top" of the cage, as at P', we have $k''x_1 = r\ (= PL)$ where x_1 is the arc OP' and as, here, the radius $P'L$ is a greater fraction (say, $\frac{3}{4}$) of the rib OP' we find $k'' = \frac{3}{4}$ and g_{22} is $3x_1/4$, a much greater value than was found in the region P farther from O.

For a (ds) triangle close to O it is easily seen that k'' will be unity; hence the chart for g_{22} (which for the screen image of the cage was a straight-sided cone) will be a cup as indicated in Fig. 32. For convenience we draw this g_{22} surface in reference to the projected image of the cage. This is satisfactory because every point in the wire cage has one definite corresponding point in the projected image. Thus if the point P' on the wire cage of Fig. 31 corresponds to P' in the projected image forming the base of the cup in Fig. 32 then the vertical line from P' up to the cup gives the value of g_{22} for the region of P'.

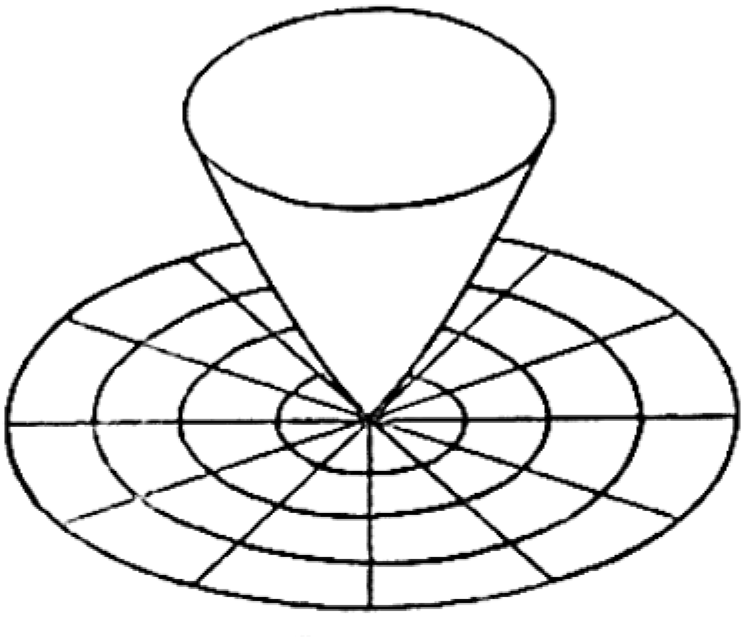

FIG. 32

Thus for the mesh-system of the actual gauze, which differs from the image on the flat screen solely in possessing a hummock, we have a cup-shaped chart whereas the screen image had a straight-sided cone. *The presence of the hummock has shown itself as curvature of the* g *chart.*

Sections of cone and cup must fit together as in Fig. 33 since the cage and its projected screen image coincide at O.

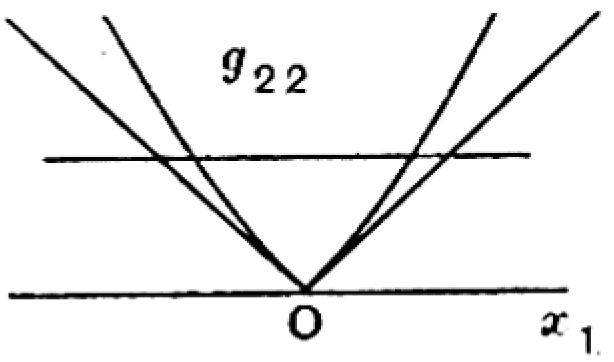

FIG. 33

As the cage is symmetrical about its axis Fig. 33 gives complete data for the values of g_{22} in every region both for the cage and for its image. The deduction of these g charts leads to the following suggestions:

1st. A rectilinear uniformly numbered mesh-system without hummocks has a horizontal g_{22} "curve," that is to say g_{22} has zero gradient.

2nd. A curvilinear mesh-system without hummocks has a g_{22} curve possessing a gradient (the gradient in the case chosen —Fig. 30—having been constant).

3rd. A curvilinear mesh-system with a hummock has a g_{22} curve which possesses a changing gradient, that is to say a gradient which itself has a gradient (in the case chosen the gradient of the g_{22} curve increases when x_1 increases).

Further tentative suggestions follow:

(*a*) When the mesh-system is uniform and rectilinear and has no hummocks the g's have no gradient (that is to say are numerical constants).

(*b*) When the mesh-system is curvilinear but has no hummocks the g's possess a gradient everywhere the same.

(*c*) When the mesh-system has hummocks the g's have a gradient of changing magnitude, that is to say they have *a gradient of a gradient.*

Suggestions thus arrived at by considering one or two special cases warrant no general conclusions but they throw some light upon the fact, which can be proved mathematically, that *the presence of a hummock in a continuum is disclosed by the gradient of the gradient of the* g'*s found for any and all of the arbitrary mesh-systems by aid of which the continuum may be explored.*

We have chosen this consideration of 2-dimensional continua as an introduction to the theory of continua of higher order because the mesh-systems laid out in continua of only two dimensions and the presence of hummocks in these continua can be visualized mentally and illustrated in line drawings. The actual continuum of space-time, to which the same principles have to be applied by the Relativist, has four dimensions and it can be shown mathematically that for such a continuum the general equation for ds^2 is

$$\begin{aligned} ds^2 = {} & g_{11}dx_1{}^2 + g_{22}dx_2{}^2 + g_{33}dx_3{}^2 + g_{44}dx_4{}^2 + g_{12}dx_1dx_2 \\ & + g_{13}dx_1dx_3 + g_{14}dx_1dx_4 + g_{23}dx_2dx_3 + g_{24}dx_2dx_4 \\ & + g_{34}dx_3dx_4 \end{aligned}$$

in which ten g's occur.

Meaning and need of the ten g's

The g's have to express the distortion of the mesh-system. We saw that in a system having two dimensions and so requiring two layers of threads and two axes, the x_1 and the x_2 axes, there were three g's, viz. g_{11}, g_{22} and g_{13}. This was because there are three factors determining the geometry of the (ds) triangles in any small region, namely, (1) the numbering of the x_1 threads, (2) the numbering of the x_2 threads and (3) the inclination of the x_1 threads to the x_2 threads. Similarly, in the general case, the number of g's is given by the sum of the number of dimensions of the continuum (the number of x's) and the number of angular inclinations that there are between pairs of sets of threads, that is to say between pairs of x's. The number of these angular inclinations is given by the number of combinations possible between the x's taken two at a time. Thus for a 3-dimensional continuum (x_1, x_2, x_3) the number of g's is

$$\{3 + (3 \times 2/2 \times 1)\} = 6$$

and for a 4-dimensional continuum the number of g's is

$$(4 + 4C_2) = (4 + 4 \times 3/2 \times 1) = 10.$$

Returning now to the problem which led us to the discussion of the g's, we remember that we set out to discover what really it was that our beetle did when he "felt out" a geodesic across a hummock. The conclusion is that by the feel of any small area ds^2 which he happened to be traversing he could detect that property of the gradient of the gradient of the g's, which characterized the hummock.

What has this to do with Relativity?

Importance of the g's to Relativity

The Relativist says force is unreal and that it must be possible to specify some observer, for any system, who will find it unnecessary to introduce "forces" into the description of that system. This means that to some specifiable observer, a set of space-time paths (e.g. those of the falling apples), which when plotted in the frame of reference (mesh-system) of *any observer whatever who regards space-time as uniform*, is found to contain bends, *shall contain no bends*.

The required specifiable observer is one who regards space-time not as uniform but as possessing hummocks, for nothing but a hummock can "straighten out" (that is to say make uniquely definable paths of) lines that are bent.

Thus the Relativist gets rid of "forces" ("bends" in the space-time paths of particles), which he has to prove can be got rid of if he is to substantiate his contention that they are "unreal," by postulating hummocks in space-time. To do this would be foolish if hummocks were themselves unreal, but if now it can be proved that the required space-time hummocks do really exist, the case for the relativity of force, including gravitational force, will have become incontestable.

In order to prove that space-time hummocks exist in the region of "matter," it must be shown that such hummocks are observed unambiguously by all observers.

The discussion of the method which Einstein has adopted in proving this demands a clear understanding of the meaning of "different observers." It must be realized that observers differ only in respect of their frames of reference—that is to say their choice of mesh-system—and that the differences between frames are expressible always in terms of the g's.

The reader may be inclined to think that the g's express *distortion* of a frame but not the motion of one frame relatively to another. He will have in mind the two frames of reference used in describing the chalk's path in the lift (p. 79). There was the lift frame, defined by the mesh-system of horizontal and vertical lines laid out by the lift-man on the wall of his lift, and there was the earth frame defined by the mesh-system of horizontal and vertical lines laid out by the ground-man on the wall of the well down which the lift fell. In respect of the former frame the chalk's path was rectilinear, in respect of the latter it was parabolic. Here we had two rectangular rectilinear mesh-systems in relative motion. The case seems hardly to be one of frame distortion. But let us consider what sort of a mesh-system the lift-man was using *from the ground-man's point of view.* This will be seen clearly if we suppose that the lift-man, while making the chalk mark with the right hand, made below it, with the left hand and at the same horizontal speed, the x axis (horizontal axis) of his own co-ordinate system. The whole wall of the lift being of gauze, the x axis as well as the chalk's path above it would both mark through upon the wall of the well, and they would make there a pair of parallel parabolas.

How would the ground-man describe what he saw?

He would say the lift-man was marking upon the *well*-wall a parabolic chalk path and was putting this into a frame of reference having a vertical rectilinear "y" axis and a *parabolic* "x" axis.

The lift-man would tell a different story; he would say that he was drawing a horizontal rectilinear chalk path and was putting it into a rectangular rectilinear frame of reference.

In other words, the ground-man would say to the lift-man: You have drawn a parabola, but as you have taken a parabolic "x" axis which you call "straight" there is no wonder that you call the parabolic path "straight" also. You are *sure* the path is straight because it is the shortest path between its beginning and its end. But my idea of straight is different from yours, and equally legitimate. From my point of view the beginning and the end of your chalk path were separated by a considerable depth of well, the line you actually drew (for here it has marked its trace upon the well-wall) was a parabola and was by no means the shortest distance between its beginning and its end. Thus not only is your "x" axis distorted, but in addition your "y" axis has been shortened. The chalk's path, which you think only a yard long, is *really* several yards long.

More about the ten g's

A complete frame of reference in space-time is a 4-dimensional mesh-system having four sets of threads and four axes, Ox_1, Ox_2, Ox_3, Ox_4.

For such systems in general a formula holds which is analogous to the general ds^2 formula already discussed for 2-dimensional systems; the formula is

$$ds^2 = g_{11}dx_1{}^2 + g_{22}dx_2{}^2 + g_{33}dx_3{}^2 \text{ etc.}$$

as already given on p. 150. It contains ten g's and by proper choice of the g's every conceivable observer's frame of reference can be represented by it.

We have said that the g's for any and every conceivable mesh-system laid over a continuum disclose, unambiguously by the gradients of their gradients, the complete character of any hummocks that may be present. But that which is expressed similarly in the equations of all conceivable mesh-systems of a continuum is something which is observed similarly by all conceivable observers, that is to say it is "real."

Hence, in order to prove that a hummock does really exist in any region in space-time—as, for instance, in the region defined by the presence of the sun or the earth, here and now—the Relativist must first define "hummock" in terms of mathematical properties of the ten g's of space-time and must then show that a series of ten or more ds^2 equations obtained, by measurements made with scales and clocks applied to "material" systems, gives a set of ten g's which do indeed possess the hummock property.

Strictly the task is an impracticable one, for the numerous measurements must be made all in *the same region* of space-time although time does not stand still. Experience has shown, however, that the mere passage of what to us seem long periods of time produces no appreciable change in gravitational fields. Our faith in this, which after all is only an experimental fact, is so complete that we regard it as an unchangeable law, the law of the conservation of mass. If time alone altered a body's gravitational field we should express the fact by saying the body's mass had changed, and change of mass has been (until recently) so contrary to experience that it has come to be thought "impossible."

Great difficulty does result, however, from the fact that the hummocks produced by the gravitational fields which characterize ordinary lumps of matter are so small that their examination, by means of equations obtained by direct measurement of time and space intervals, is impracticable; it is impracticable, for instance, to find the deviation from π produced by putting a lump of lead at the centre of a circle: measures of length cannot be made to the degree of accuracy required to show up the differences in the g's, which express the hummock produced by the gravitational field of the lump of lead.

The claim made by Relativity, that the reality underlying a gravitational field is a "hummock" in space-time, is, in effect, a claim that the law of gravitation is a law of hummocks, and must be derivable from the principle of Relativity without the introduction of any independent hypotheses such as Newton had to make.

Einstein has boldly accepted this position and has derived from the g equations a "law of gravitation" which proves to be more general and more accurate than Newton's.

It is not surprising that, in deriving his new law, Einstein had to make some tentative assumptions. Experimental data as to the exact character of the hummocks characterizing "matter" being meagre and inexact, it was necessary to make some arbitrary assumptions in regard to these hummocks and see how these assumptions accorded with the known facts about gravitational fields.

Tentatively Einstein assumed that "space-time in the region of, but not containing matter, is curved only in the first degree." This statement is not possible of definition except in mathematical terms. By rough analogy we may approach some sort of a physical conception of what is meant, but that is all. The expression "4-dimensional region containing matter" has a meaning which cannot be visualized, or, in the ordinary sense, conceived, and the difficulty of comprehending a "hummock" or "curvature" in space-time is insurmountable by "common-sense." Some suggestion of what is meant by "curved in the first degree," as applied to 4-dimensional space-time, may be found by analogy. A 1-dimensional continuum (a line) may be curved either into a flat curve (like that of a watch spring) occupying only two dimensions, or into a helical curve (like that of a cork-screw) occupying three dimensions; by analogy, therefore, a 2-dimensional continuum (a surface) may be curved not only into the third spatial dimension to form the ordinary comprehensible sort of hummock, but also in some more complex way which cannot be visualized, and a 3-dimensional continuum has mathematically possible curvatures of still greater complexity, while, lastly, our 4-dimensional space-time continuum has still more formidable possibilities of complexity of curvature. Einstein hoped and, therefore, tentatively assumed, that, for all practical purposes of description, space-time curvature (which is the defining characteristic of a hummock) of the simplest among the formidable variety of "degrees of curvature" of which a 4-dimensional continuum is capable, might suffice.

The law of gravitation which Einstein has formulated on the basis of this assumption is found consistent with all the known phenomena of mechanics and of gravitation. This law differs greatly from Newton's though for all ordinary phenomena the difference is quantitatively very small—usually, in fact, negligible.

Newton's law of gravitation is a special case of Einstein's

Such cases as this occur often in practice. For instance: a well-known law of physics, which happens also to have been invented by Newton, relating to the rate at which hot bodies cool, reads: "The rate at which a hot body loses temperature is directly proportional to the excess of the body's temperature above that of the surrounding medium," or, more briefly, "the hotter a thing is the quicker it cools." This law works very well in most ordinary cases but it is inaccurate and is only a special case of the more general and complex law of Stefan which reads: "The rate at which a body loses energy by radiation to the surrounding medium is proportional to the excess of the 4th power of the absolute temperature of the body above the 4th power of the absolute temperature of the medium." The two laws differ greatly, but in their application to many ordinary cases the difference is, quantitatively, small. Newton's law of cooling is about right for cases in which the excess of temperature is small while the more general law of Stefan applies over a practically unlimited range. Thus Newton's law may be regarded as a special case of Stefan's. Similarly in regard to Newton's law of gravitation in relation to Einstein's; the former applies accurately over a limited range of conditions (the limits of which are, however, remarkably wide), while the latter promises to have a practically unlimited range.

CHAPTER VIII

The laws of space-time hummocks

BEFORE trying to explain something of the way in which Einstein's law of gravitation was derived and how it was checked by actual experiment, we have to note two further points in regard to the ds^2 method.

From the point of view of Relativity the law of gravitation is the rule (or the closest approach that can be found to such a rule) which nature has followed in providing space-time with hummocks. Nature has made these hummocks; they are "absolute" and we have to take them as we find them. The "why?" and "how?" of these hummocks are, at present, meaningless questions, the problem of science is simply to explore and describe them.

In accordance with the beetle analogy, hummocks must be explorable, completely, by means of measurements of groups of infinitely small intervals. That which is happening in any little element of space-time "area" ds^2 defines completely that which is happening in the immediate neighbourhood, and so on throughout the "world." From this it follows that nothing happens in any element ds^2 without affecting the (ds^2)'s around it, and as the changes in these latter elements must effect changes in surrounding elements, it follows that the world as a whole is a continuous organism of infinite sensibility and that the notion of action at a distance is unnecessary in the description of phenomena.

How then can the Relativist use the facts of science in the form in which they always are given—namely as statements in regard to particles?—for atomicity is opposed to continuity.

Relativity and matter

The answer is that the Relativist's notion of continuity is applicable only to regions of space-time which are explorable by measuring instruments. Obviously this is so since the notion of the continuity of the world is based upon the practical success with which the ds^2 equations have been found applicable to experimental phenomena, that is to say to events capable of description in terms of measurements. Ultimate particles of matter can have no structure: the idea of measuring them or their contents is meaningless; hence it is meaningless to make postulates about ds^2 equations inside ultimate particles, and meaningless to claim continuity for the regions in space-time defined by such particles. At present electrons are playing the part of ultimate particles, and Relativity has nothing to say about the region inside an electron.

Coincidences in space-time are essential factors in reality

Our discussion of the ds^2 equations seems to have put into the background the statement that Relativity finds reality in nothing except in space-time coincidences and their ordering: hummocks seem to have eclipsed everything.

The position may be made clear by remembering that "hummocks" is the name given to certain relations of the g's of ds^2 equations. We need only show, therefore, that the methods of ds^2 equations are consistent with what has been said about space-time coincidences. Such equations express relations of intervals measured in space-time, and nothing more. But to measure an interval—a space interval or a time interval—means to record the coincidence of a pointer with a mark on a scale (or it means the equivalent of this) and nothing more. Hence ds^2 equations, and so also hummocks, concern nothing except space-time coincidences and their relations—that is to say their "ordering."

Hummocks are expressed by those properties of the g's (especially the gradients of their gradients) which are "invariant," that is to say which remain unchanged in the ds^2 equations no matter what sort of mesh-system may have been used as a basis for the ds triangles; and we have seen that a change of mesh-system is equivalent to a distortion of the continuum—but always to a sort of distortion which neither produces nor affects hummocks. The only factors in the system of world-lines embedded in a continuum which can remain unaffected by arbitrary distortions are intersections and their "ordering." Hence it is these intersections and their ordering which hummocks express, and Relativity regards space-time hummocks as the world's only scientific reality.

Is it, however, the hummocks in space-time that are real or is it space-time with its hummocks, or is it not rather the experimental measurement factors to express whose relations the ideas of space-time and of hummocks have been invented?

Such questions are difficult to answer.

Measurements are observations of coincidences in space and time, and it seems allowable to state, without attempt at definition, that all men know and are in agreement as to what a space-time coincidence is. The experience of any individual observer is made up of space-time coincidences and consists in observation of their relations or mode of ordering in space and in time, and each observer knows to his own satisfaction what is the meaning of such ordering. Science has lately shown that the notions which different individual observers have of the space ordering and of the time ordering of one and the same group of coincidences are not necessarily either consistent or reconcilable, and to overcome this difficulty—which arises out of the unwarrantable assumption which individual observers make about an "ether" of space and an independent uniformly flowing time—the Theory of Relativity has been invented, which proposes that the "reality" in the ordering of coincidences is an ordering neither in space nor in time (considered independently) but in a continuum of space and time.

Perhaps therefore the difficult questions may be answered thus:

The only reality—to the individual—is individual experience and this belongs to and can be realized only by the individual, but if I assume that there exist other individuals like myself, and that to all of us there come similar experiences out of an external world, then the reality of this external world consists of indefinable but universally recognized absolute things called "space-time coincidences" whose "space-time ordering," but nothing more, is observable and similarly observable by all observers. Space-time ordering cannot be defined except by saying that it is the only factor in experience in regard to which all observers agree. In other words, space-time coincidences and their ordering compose the world of the universal observer.

It is impossible to visualize a 4-dimensional continuum or a mesh-system filling it. Always in discussing it we have to refer to some analogous case of a continuum of fewer dimensions.

The position of a point in the space-time continuum requires four co-ordinates (x_1, x_2, x_3, x_4) for its specification, these co-ordinates being defined by numbers upon the four sets of threads of a mesh-system whose axes of reference are four threads passing through a common "point" called the origin of co-ordinates (this "point" is a thing we cannot visualize in a 4-dimensional continuum). If such a mesh-system is to be regarded as rectangular and rectilinear, then the four axes of co-ordinates are said to be "mutually perpendicular."

As space is 3-dimensional we are unable to visualize four mutually perpendicular lines.

A hummock can be visualized in a 2-dimensional continuum. One example of a 2-dimensional continuum is a Euclidean plane surface, say, the surface of a flat sheet of copper. A hummock may be produced in such a surface by beating the sheet with a hammer until it becomes bowl-shaped. Such a hummock is a hummock in space. The surface is now no longer Euclidean, for the circumference of a circle described upon it is no longer π times the diameter, nor is the sum of the three angles of a triangle, having "straight lines" (that is to say "shortest paths" or uniquely definable paths) for its sides, equal to two right angles.

According to Relativity the effect of a gravitating mass is to distort the space-time regions in which it lies. This distortion consists partly of a space distortion, in an endeavour to visualize which we define a 2-dimensional sheet of space—the surface, for instance, of a drawing board—and place a lump of lead at its centre. Presumably the distortion of the region will make a "hummock" upon the surface of the drawing board, the presence of which may be detectable by describing a circle about the lump of lead and measuring the ratio of the circumference to the diameter of the circle.

Let *PQR* (Fig. 34) be such a circle, described symmetrically within a polar mesh-system.

Then for this system while undistorted—that is to say before the lump of lead is placed at *O*—we have

$$ds^2 = dx_1^2 + x_1 dx_2^2,$$

while, with the lead at *O* the equation would have to be expressed in the general form,

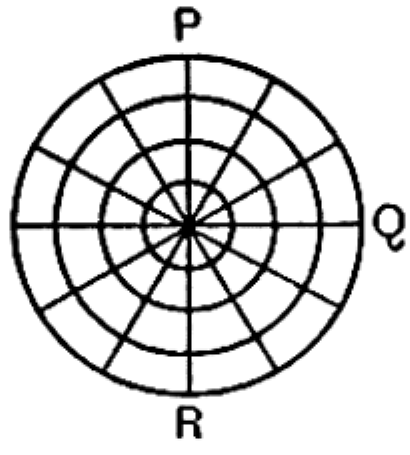

FIG. 34

$$ds^2 = g_{11} dx_1^2 + g_{12} dx_1 dx_2 + g_{22} dx_2^2,$$

and the *g*'s would have to be found by experiment. We have pointed out already, however, that the measurements required to show the divergence of the *g*'s from the values they have in the preceding equation (namely, $g_{11} = 1$, $g_{12} = 0$ and $g_{22} = x_1$) are impracticable because of the smallness of the distortion.

The minuteness of the distortions which had to be explored made it necessary to base the law of gravitation upon measurements which were not directly measurements of *ds* triangles.

Einstein's law of gravitation

Einstein believed that Newton's law of gravitation was wrong because it was, at any rate, ambiguous. It stated, for instance, that gravitational effects depended upon the product of the "masses" of the bodies concerned, yet, not only Einstein but also the earlier Relativists had long known that "mass" is a relative quantity, and the mass of a body ambiguous.

It may be objected that Newtonian mass cannot change since Newton defined it as an unchanging characteristic of a body. If this point of view be taken it then becomes necessary to say that Newton's law of conservation of momentum is at fault. The Relativist finds it simpler to retain the law of conservation and to re-define mass so that momentum is conserved. It would be possible so to define "mass" that Newton's law of gravitation should be strictly correct, but the definition thus required for mass would be complex and impracticable.

We shall see later that the law of conservation of momentum can be retained only by defining mass as a quantity which increases with velocity.

Newton's law of gravitation is true in so far as it works, and it does so with very remarkable accuracy over an exceedingly great range of conditions. Einstein knew that for all ordinary mechanical phenomena, and for nearly all known astronomical ones, his new law would have to be practically indistinguishable from Newton's.

How had Newton erred?

How had Newton erred?

It is misleading to say that he had erred in introducing the idea of "force" into his description of gravitational phenomena. The thing observed was the path of a particle. This path may be described as "bent" because of a "force," or "straight" because of a "hummock."

Newton erred in supposing that our world was so constructed that the bend in a particle's path could be analysed legitimately into a space bend and a time bend.

The meaning of this may be explained by reference to the path of a bullet fired horizontally from the top of a cliff. In Fig. 35, *OP* represents a vertical cliff and *Ox* the sea's surface, while *PNQ* is the space path of a bullet fired horizontally from *P*. According to Newtonian description the bend in *PNQ* is due to the "force" of gravitation. The bullet is said to follow a parabolic path in space.

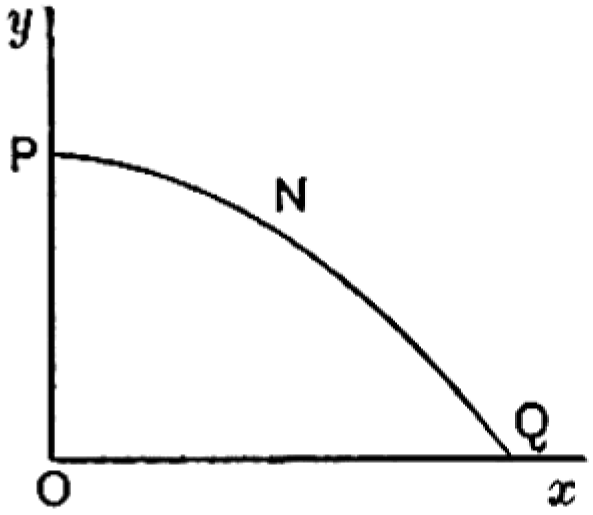

FIG. 35

Fig. 36 is another diagram referring to the same bullet's path. Here the horizontal axis represents time. Starting at *P* the bullet falls from *P* to *O* in the time *Ot* with ever increasing *vertical* velocity downwards. The "bend" in this diagram occurs in "time."

Thus according to the Newtonian method it is regarded as legitimate to record the space bend and the time bend independently.

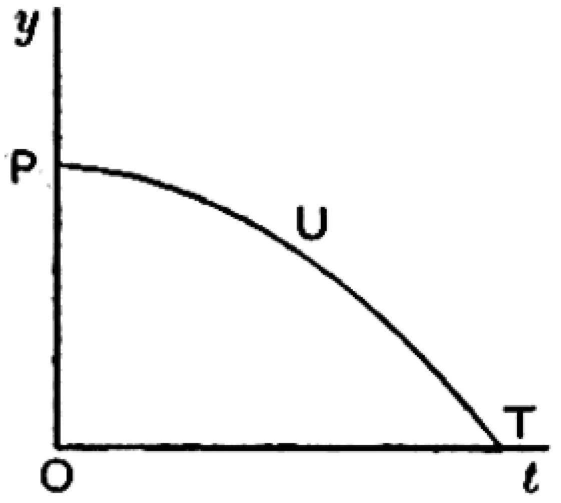

FIG. 36

According to Relativity this procedure is not legitimate but gives of the phenomenon a description which is not "true." The Newtonian method fails to take proper account of the fact that the bullet's path is in space-time and that no analysis of space-time into separate space and time can be true over an extended region. Such analysis might be legitimate if we knew how to define "simultaneity in time at two different places." Newton assumed that we knew how to do this; Relativity faces the fact that we do not know how, and prefers to say that until we have learned how we had better regard the very idea of simultaneity over an extended region as meaningless.

In order to throw light upon this important point we consider Figs. 35 and 36 in more detail and with reference to the nature of the experimental procedure required in finding data for these diagrams.

We neglect experimental difficulties and assume, for instance, that the bullet is a hollow ball containing an observer provided with a pair of telescopes so arranged that the one remains always horizontal and the other always vertical during the flight of the ball. We assume further that the vertical side of the cliff has been marked with a scale of feet and that, somehow, a horizontal scale of feet has been marked also upon the surface of the sea. All that the observer has then to do is to record a series of successive simultaneous pairs of readings of his two telescopes directed, as these are, upon the y and x scale respectively. In Fig. 37, B represents the ball, h and v representing the horizontal and vertical telescopes respectively.

Each pair of readings gives a value of y and a value of x corresponding to one particular point on the curve.

It seems as though no unwarranted assumption has been

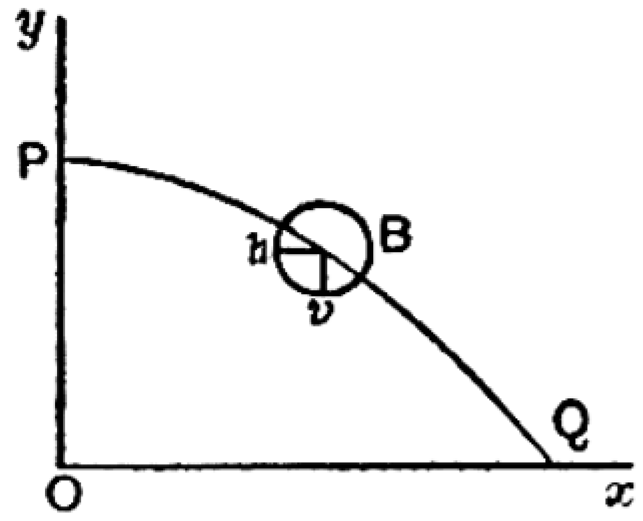

FIG. 37

made, and certainly none in regard to time, since time has not entered into the measurements. But in saying this are we not overlooking the time taken for the light to travel from the cliff and from the sea to the telescopes? In any practical case such time periods are too short for any detectable change in the ball's position, but it is the principle of the experiment which we are discussing, not its practical accuracy.

It may be replied that the velocity of light is known and we can make, in principle, exact allowance for the times involved and so can make the curve perfectly accurate. But this is to overlook the fact that the so-called "velocity of light" is its velocity relative to the ether which transmits it and that no observer either knows or can find out his own velocity relative to the ether.

The time element is in fact not absent from the experiment of Fig. 37 although only space measurements appear to be involved, and the Newtonian method cannot take proper account of this time element. The problem is a space-time problem and, for unlimited accuracy it must be dealt with as such.

What of the Newtonian "time" curve of Fig. 36? To simplify the consideration of the problem we imagine a rigid vertical wall erected close to the trajectory of the ball and forming a background to this trajectory. Upon this background (exactly represented by the plane surface of the paper in Fig. 37) are drawn horizontal lines one foot apart. The observer in the ball carries a watch and all he has to do in order to find data for Fig. 36 is to record the readings of his watch at the instants at which the ball passes the successive divisions (horizontal lines) marked upon the background. As the ball's path is very close to the wall there is no ambiguity about the readings for each reading is the observation of a coincidence in space and time of the ball, a horizontal line, the hand of the watch and a division on the watch dial, these factors all coming into collision at once for each reading. The observer thus obtains successive pairs of (y, t) values, each pair defining one point on the space-time curve of Fig. 36.

Here at any rate the Relativist can hardly drag in his favourite bogey the velocity of light. No, but this time he tells you he cannot accept the readings of the observer's watch because it was a watch constructed and calibrated on *terra firma*—that is to say in a gravitational field—and the observer used it inside a freely falling ball in which there is no gravitational field, and (as will be shown later) a clock's rate of going is affected by gravitational fields.

In this experiment, therefore, error has crept in on account of the unwarranted assumption of Newtonian mechanics, that "time flows uniformly." We do not really know how time flows, nor does it matter, but we can prove experimentally that clocks are affected by gravitational fields.

In spite of such difficulties it was necessary for Einstein to evolve a law of gravitation, guided by experimental results expressed in the customary Newtonian way, and by his own assumption that space-time is curved only in the first degree.

Einstein's problem

The world consists of particles in motion, and to describe the world means to describe the space-time paths of these particles. The description must be "true" and so must be unambiguous. To describe a path either as straight or as bent is ambiguous except to one particular individual observer, and science seeks a description which shall be unambiguous even to the universal observer (who embodies all conceivable individual observers).

Only one method has been found by which this can be done; it is the method of Relativity, in which a path is not in space or in time but in space-time and is measured in units of "interval" (a mathematical compound of space and time).

The universal observer finds one and only one space-time path capable of unambiguous description—which, therefore, consistently with our definition of reality, must be the particle's "real" and actual path—it is the path of "*greatest possible interval length.*"

Greatest possible interval length has a definite mathematical meaning in terms of space-time measurements, analogous to that of "shortest possible distance" in Newtonian space.

Einstein has predicted that all free particles will be found to follow such paths; that is to say he has proposed a new law of motion which runs: "Every world-line has the maximum possible interval length," or "Every particle goes (*extends* might be a less inappropriate word since the idea of motion has no counterpart in the space-time world) from 'point' to 'point' by the longest possible path." This is the law which the universal observer finds to be obeyed in all natural phenomena.

How does he test nature in regard to its obedience to this law?

It has been explained that the only "real" measurements are those of space-time intervals. The grounds for this will be explained in more detail in a chapter on the Special Theory of Relativity.

The "interval" ds between a "point" (x_1, x_2, x_3, x_4) in space-time and another "point" very close to it is represented, in *any* mesh-system, by the equation

$$\begin{aligned} ds^2 = {} & g_{11}dx_1^2 + g_{22}dx_2^2 + g_{33}dx_3^2 + g_{44}dx_4^2 + g_{12}dx_1dx_2 \\ & + g_{13}dx_1dx_3 + g_{14}dx_1dx_4 + g_{23}dx_2dx_3 + g_{24}dx_2dx_4 \\ & + g_{34}dx_3dx_4 \end{aligned}$$

in which dx_1, dx_2 and dx_3 represent the elements of space and dx_4 represents $(\sqrt{-1})$ times the element of "time" separating the two points. The quantity $(\sqrt{-1})$ has to be introduced, that is to say dx_4^2 has to be made to correspond with $(-dt^2)$ instead of dt^2 (as analogy would demand), in order that ds shall be a quantity "similarly observed by all observers." It is because of this $(\sqrt{-1})$, which nature demands, that the analogy between ordinary space and space-time is of only limited application.

The values of the g's depend upon the mesh-system chosen, as well as upon the presence of hummocks, but there are certain mathematical properties of the g's, which occur independently of the choice of mesh-system, and these disclose the hummocks. For this reason the hummocks, being absolute characteristics, are said to determine the nature of the space-time of the region concerned.

To say that the choice of mesh-system is without effect upon either the (ds)'s or the g's, is equivalent to saying that any individual observer may use his own clock and scale (properly standardized in the Newtonian fashion) in measuring the (dx)'s and may be sure of finding out the true value of ds and the true character of the space-time. All he has to do is to use the ds^2 formula.

Einstein knew that Newton's laws were certainly *nearly* right; he knew, therefore, with considerable accuracy, how a particle of matter would behave in a gravitational field, as, for instance, the earth in the sun's field. This means that Einstein knew how to plot the path of such a particle in Newtonian space and time; and, as elements of "interval" are defined in terms of elements of Newtonian space and time, it was possible to write out a number of ds^2 equations (with the g's as the "unknowns"), for any given region in, say, the sun's field. In order to do this it was necessary to choose, arbitrarily, some mesh-system, as, for instance, in the case of a planet moving round the sun in a plane orbit, a polar system with the sun as origin. With this choice of mesh-system the ds^2 equation would be written without the terms involving x_3, since, in taking a particle moving in a plane orbit only two dimensions of space are concerned.

By solving such ds^2 equations, leading to tentative values of the ten g's in neighbouring regions, and adapting the results to his own condition that the g's should disclose only curvature of the "first degree," as well as to the condition that in regions spatially remote from the origin the g's should assume values corresponding to those of "flat" space-time—values whose gradient possessed no gradient—Einstein found the following values for the g's:

$$g_{11} = +\ 1/\gamma, \quad g_{22} = +\ x_1^2, \quad g_{44} = +\ \gamma,$$
$$g_{12} = 0, \qquad g_{14} = 0, \qquad g_{24} = 0.$$

Thus the ds^2 equation is,

$$ds^2 = (+\ 1/\gamma)\, dx_1^2 + x_1^2 dx_2^2 + \gamma\, dx_4^2,$$

or, writing r for x_1, θ for x_2 and $(\sqrt{-1})\, t$ for x_4,

Equation of Einstein's law of gravitation

$$ds^2 = (1/\gamma)\, dr^2 + r^2 d\theta^2 - \gamma\, dt^2.$$

The symbol γ stands for $(1 - 2m/r)$ where m may be any constant, or, with proper choice of units, represents the gravitational mass of the gravitating body.

The problem of finding these g's was long and complicated. Strict adherence to the quantities deduced from Newton's laws led to values of the gradients of the gradients of the g's which were *not unaffected* by change of mesh-system. This showed Newton's law to be wrong, for the alternative conclusion, namely that hummocks in space-time are not unambiguous and so are not real, was contrary to Relativity.

The ds^2 equations had to be modified in such a way that the gradients of the gradients of the g's were unaffected by change of mesh-system. To do this required the application of the complicated and difficult mathematical "theory of tensors" which, fortunately, had been worked out long ago by such mathematicians as Christoffel and Riemann, who had no idea that the theory might some day find practical application.

The reasonableness of the "γ" equation may be seen by inspection.

The quantity $(1 - 2m/r)$ approaches the value unity when r becomes great, hence, in regions remote from the gravitating mass, the ds^2 equation reads:

$$ds^2 = dr^2 + r^2 d\theta^2 - dt^2$$

or

$$ds^2 = dx_1^2 + x_1 dx_2^2 + dx_4^2.$$

This is the form corresponding to the equation

$$ds^2 = dx_1{}^2 + x_1 dx_2{}^2$$

which we found on p. 144 corresponded to ordinary flat space; it is a form in which the g's (viz. $g_{11} = 1$, $g_{22} = x_1$, $g_{44} = 1$) have no gradients of gradients. g_{11} and g_{44} being numerical constants have no gradients which might possess gradients, while g_{22} has a gradient (namely, g_{22} varies directly with x_1) which is constant (see Fig. 30 on p. 145) and so possesses no gradient.

Thus the "γ" equation, when r is great, corresponds to space-time without hummocks, that is to say to a region in which no permanent gravitational field is appreciable—a region remote from gravitating mass.

Again, on account of the well-known symmetry of the gravitational field, the force at a given distance from the mass is the same for all values of θ, hence it is reasonable to find that none of the g's in the equation for the plane orbit is a function of θ; while, lastly, the polar form of the mesh-system accounts for the value r^2 for g_{22}.

CHAPTER IX

The slowing of clocks and the shortening of rods in a gravitational field

EVERYONE has been struck by the statement that, according to Relativity, the going of clocks and the dimensions of rigid bodies are affected by gravitational fields. This matter will be dealt with from simple points of view in later chapters but it is appropriate here to show the way in which these strange effects are demanded by the "γ" equation.

It is necessary first to realize that statements about the going of clocks and the lengths of bodies are meaningless to Relativity. There is no such thing as an absolute time interval nor an absolute length interval, and to speak thus of clocks and rods is to confess oneself a biassed and peculiar individual—a Newtonian observer. The fact has to be faced, however, that each one of us is such an observer and must remain such: the doctrines of the universal observer leave us cold.

If now we suppose that the "γ" equation relates, say, to the earth's gravitational field, and the question is raised of the effect of that field upon clocks and rods we shall have to put ourselves with our own standard clock and rod in some region right outside the earth's field and so, looking on from afar, we shall judge of the happenings in the earth's field. We are hardly likely to detect a slowing of clocks or a contraction of rods if we ourselves are near the clocks and rod so that our own heart beats are slowed just as the clocks are and our own height contracts just as our measuring rod does—in fact, by the Principle of Equivalence (see p. 193) on which Einstein bases his General Theory, every observer is justified in maintaining that his own peculiar clock and rod are "true" standards.

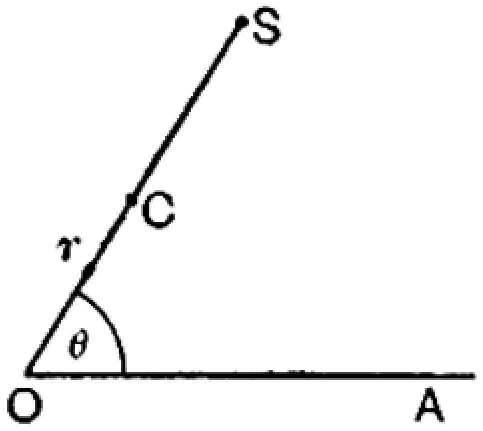

FIG. 38.

If Einstein's "γ" equation is right what shall we observe when we look in from afar into the earth's field? The "γ" equation tells us how to measure a small "interval" ds. We place ourselves at S (Fig. 38), a great distance from earth (O) and we observe a clock at C distant r from O and stationary in space.

By remaining just where it is in space for a time dt the clock describes the element ds of space-time interval given by the equation

$$ds^2 = (-1/\gamma)\, dr^2 - r^2 d\theta^2 + \gamma dt^2$$

in which $dr = 0$ and $d\theta = 0$, because to remain stationary during the time dt means to make no change in r or in θ.

Now by the Principle of Equivalence the man at C with his clock has a right to say *his own clock is going properly and his own rod is a true standard*, and therefore to use for his own immediate vicinity the equation for "flat" space-time, namely

$$ds^2 = -dr^2 - r^2 d\theta^2 + dt^2,$$

and to say that since he and his clock remain together and at rest during time dt

$$ds^2 = -0 - 0 + dt^2,$$

$$\therefore\ ds = dt,$$

that is to say C is warranted in saying that, under these conditions, his own clock reads elements of "interval" ds.

But "interval" is the one thing of which all observers including both C and S make similar measurement, hence that same interval which C calls $ds =$ say, 5; S must call also $ds = 5$; but whereas C writes

$$ds = 5 = dt = 5 \text{ seconds},$$

S will write $$ds = 5 = (\sqrt{\gamma})\, dt,$$

so that if, for instance, the value of $\sqrt{\gamma}$ were, say, 5/9, S would say $dt = 5 \times 9/5 = 9$ seconds and would legitimately conclude that 9 seconds had elapsed while C's clock was registering only 5 seconds and therefore that the effect of the earth's field upon C's clock was to make it go slowly.

(It is hardly necessary to mention that the actual value of γ in the earth's field is so little different from unity that the actual slowing of clocks is infinitesimal.)

Next we consider C's measuring rod. We suppose that in practically no time the observer C, keeping his measuring rod along a given radius of the earth (and so keeping θ constant and making $d\theta = 0$), measures the height of a vertical post and finds it 2 feet.

By the Principle of Equivalence C is warranted in saying that his own rod is of true standard length and his own clock is ticking true seconds, so he writes the equation for "flat" space-time,

$$ds^2 = dr^2 + r^2 d\theta^2 - dt^2$$

(the change of sign is made for convenience in this equation; its legitimacy is discussed on p. 180), and as dt and $d\theta$ were both nought the equation becomes,

$$ds^2 = dr^2,$$

whence

$$ds = dr = 2 \text{ feet.}$$

S has watched the proceeding from afar and knows that, for the region in which C dwells, the "γ" equation applies, namely:

$$ds^2 = (1/\gamma)\, dr^2 + r^2 d\theta^2 - \gamma\, dt^2.$$

In measuring the height of the post C had taken no time (and as S and C are at rest relatively to each other there can be no ambiguity about the meaning of "no time"), so that dt was nought, and had made no change in θ, so that $d\theta$ also was nought; hence S writes

$$ds^2 = (1/\gamma)\, dr^2 \text{ and } ds = (\sqrt{1/\gamma})\, dr.$$

As ds is a quantity found numerically the same by all observers (equipped, as they are assumed to be, with rod, and clocks which when placed at rest near standards, also at rest, are in agreement) S must find for ds the same value 2 which C found; but S writes

$$ds = 2 = (\sqrt{1/\gamma})\, dr$$

and if $\sqrt{1/\gamma}$ be equal, say, to 9/5, this means that, according to S',

$$ds = 2 = (9/5)\, dr$$

or dr, the length of the post, is $(2 \div 9/5) = (10/9)$ feet.

Thus it takes 2 of C's feet to make $1\frac{1}{9}$ of S's feet or C's feet are only about half as long as S's and S says C's rod has contracted—on account of the earth's gravitational field.

It is interesting to note that while the "γ" equation demands a contraction in a rod laid along the gravitational "lines of force," it demands no change in a rod laid transversely. For suppose C shall be measuring the distance between two points 2 feet apart, both on the earth's equator, C considers that the equation

$$ds^2 = dr^2 + r^2 d\theta^2 - dt^2$$

holds. He laid his rod practically along the equator, so that during the measurement no change occurred in r and dr was nought. He made the measurement in *no time*, hence dt was nought and he wrote

$$ds^2 = r^2 d\theta^2 = R^2 (l/R)^2 = l^2,$$

where l is the distance he finds between the points and R is earth's radius (so that $d\theta = l/R$ in circular measure). Thus C finds $ds = l =$ say, 6 feet.

S has watched the proceeding and, in applying the equation

$$ds^2 = (1/\gamma)\, dr^2 + r^2 d\theta^2 - \gamma\, dt^2,$$

he also sets $dr = 0$ and $dt = 0$, and writes

$$ds^2 = r^2 d\theta^2 = R^2 (l/R)^2.$$

For ds he, like C and all conceivable observers, must obtain the value 6; hence S writes

$$6 = ds = l = 6 \text{ feet}.$$

Thus C and S agree as to transverse measurements, that is to say the transverse dimensions of bodies are unaffected by a gravitational field.

These examples show what is meant when it is said that, according to Relativity, the relative sizes of the units of length and of time used by different observers depend upon the relative nature of the space-time they inhabit. There are no absolute units of length or of time, the meaning is that if I, living here in what I call locally undistorted space-time, look into the region inhabited by other observers, then if that region contains gravitational fields (characterized, as they must be, by hummocks) I shall conclude that those observers are using "wrong" units of length and of time.

It is well here to deal with an apparent inconsistency. The important property of hummocks in space-time is that they are "real," that is to say they are unambiguously perceived by all observers. Thus if a region in space-time contains hummocks, no observer, not even one living on the "top" of a hummock can fail to observe the hummock. How can this be made consistent with the Principle of Equivalence which allows any and every observer to use the equation for "flat" space-time, that is to say allows him to assume that his own clock and rod are "true" standards?

The difficulty may be made clear by an analogy. Although there are in space-time absolute hummocks (like hills, by analogy) there is no absolute normal or zero level analogous to sea-level, and no observer is better justified than another in maintaining that he himself lives at normal level in space-time.

Again, on the top of a hill or at any point on a steep slope, a small area ds^2 is indistinguishable from a "flat" surface. The same holds in space-time. Thus *any* observer (C) who chooses to look for hummocks in the space-time he inhabits, using, as he must, the method of measuring up a number of ds triangles, may, if he likes, assume that, for the triangle he himself inhabits, the formula for "flat" space-time holds, namely (in a polar mesh-system)

$$ds^2 = dr^2 + r^2 d\theta^2 - dt^2.$$

This will prove to be a particular case of the general equation

$$ds^2 = g_{11} dr^2 + g_{22} d\theta^2 + g_{44} dt^2$$

for which he will find the value of the g's by his measurement of numerous triangles.

But how can we reconcile this with the statement made by some distant observer (S) that, for the very same small region for which C wrote the "flat" equation, the ds^2 equation should be

$$ds^2 = (1/\gamma)\, dr^2 + r^2 d\theta^2 - \gamma\, dt^2$$

where γ is *not unity*?

The explanation is that, as pointed out on p. 146, the actual numerical values of the g's depend upon the units in which the sides of the ds triangles are measured. C and S use different units of length and of time, that is why they assign, for the very same local region, different sets of numerical values to the g's, C saying, for instance, that $g_{11} = 1$ and $g_{44} = -1$ while S says that $g_{11} = (1/\gamma)$ and $g_{44} = -\gamma$.

The important fact is this, that, in spite of their different units (as also in spite of a possibly different choice of mesh-system), both C and S will find sets of g's whose gradients of their gradients are the same. Thus both C and S discover the same hummocks.

To speak of the effect of gravitational fields upon the units of length and of time employed by different observers is a favourite method of giving a "practical explanation" of Relativity, but it has many shortcomings and because it is an attempt to explain the space-time world by the use of an inadequate imagery, it is apt to lead to confusion and error. As, however, it is a favourite method, something more must be said about it. The method employs what may be called "interval maps," which are supposed to demonstrate the interval relations of space-time (or a section of it) in much the same way that the ordinary scale-map of a country demonstrates the distance relations of space (or a surface in space).

London's "inner circle" railway may be shown *diagrammatically* as, say, a circle or an ellipse, with the stations evenly spaced around the periphery, but in a true *map* the "circle" has an irregular shape and the stations are spaced irregularly. The map may be called true provided it gives distances in their true proportions.

A true map of events in space-time must be a diagram in which the relations of events are shown in their "true" proportions. Spaces, and time intervals are relative; the "true" interval between two point-events (a point-event being the name used for a point in space at an instant in time) is neither a space interval nor a time interval but is an element of combined space and time, and it is defined as a measure-quantity whose numerical value is found the same by all conceivable observers.

In "flat" space-time the interval ds is defined by the equation

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2,$$

where dx_4^2 stands for $(-dt)^2$, and in any sort of space-time known to occur in nature an observer may employ this equation for his own immediate local region ds^2. In doing so he merely is postulating that his own clock and rod shall be "true" standards. If he live in a hummocked region he will find that observers distant from him are using wrong standards. Before going out into those distant regions it is true that they all came and compared their clocks and rods with his own and found them in agreement but they have now gone into distorted regions in which the going of clocks and the lengths of rods have been affected.

In a map of an extended region this change of units has to be shown by appropriate changes in the spacings of the mesh-lines defining equal spaces and equal times in the map.

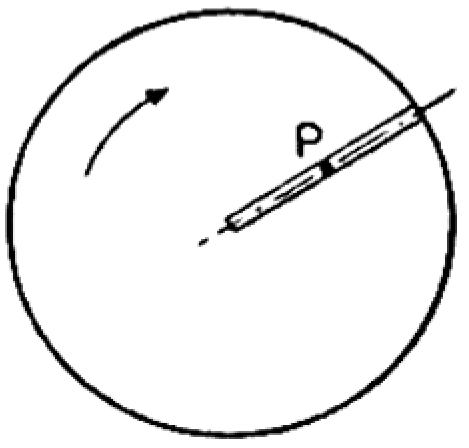

FIG. 39

We will consider an observer located, far away from permanent gravitational fields, at the centre of a rotating disc. This observer sets out to record upon a flat chart the space-time path of a particle P (Fig. 39) moving along a tube fixed radially upon the rotating disc. The particle flies outward with ever increasing velocity (its radial acceleration being $r\omega^2$ where r is the radial distance OP and ω is the angular speed of rotation of the disc). The observer, O, probably would use rectangular co-ordinates and would plot such a curve as that shown in Fig. 40. This diagram is

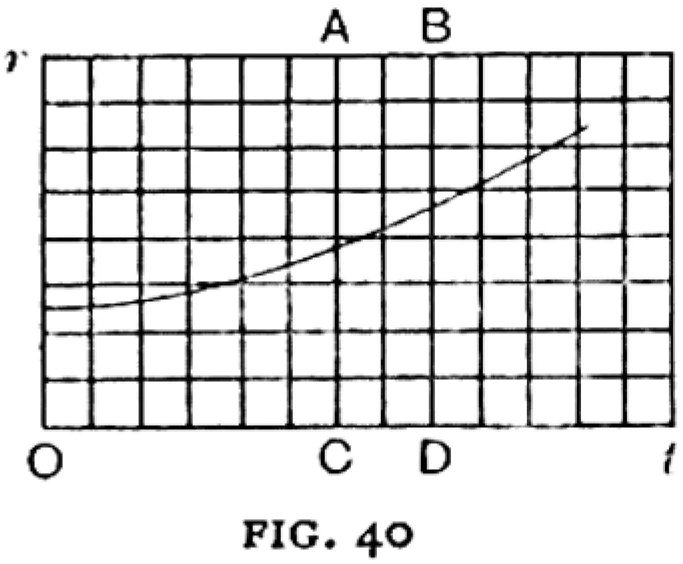

FIG. 40

not a true map. It is not easy to show just how Fig. 40 will have to be modified in order to make it "true," but one step in the required modification is fairly obvious.

Fig. 40 indicates that the unit of time is the same on all parts of the disc. We know, however, that in regions far from the centre of the disc the gravitational (centrifugal) field is strong and clocks will go slowly—as seen by the observer O. Let us fix a clock at radius CA (Fig. 40). In the disc frame of reference this clock is stationary and an observer close to the clock A would write

$$ds^2 = -dr^2 - r^2 d\theta^2 + dt^2$$

in which $dr = 0$ and $d\theta = 0$, and would rightly conclude that the clock was ticking units of interval (since $ds = dt$). But the clock A is going slowly, that is to say between two successive ticks *more than one unit of time passes*; or, *in one unit of interval there is more than one unit of time*. But in a "map" one unit of interval is to be represented by, say, one inch or some other definite paper distance; hence to say that in the region AB more than one unit of time must be put into unit interval is to say that into each inch of paper at AB more time divisions (warp threads) must be crowded than into one inch at CD.

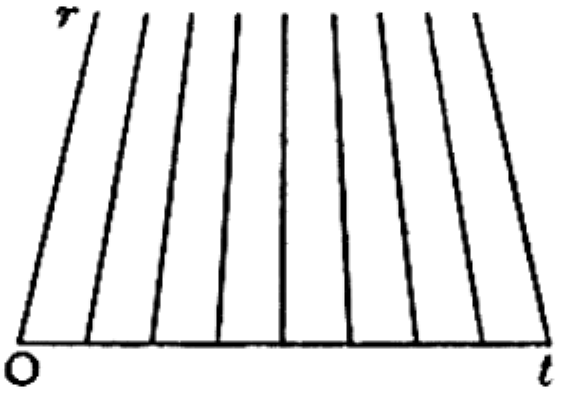

FIG. 41

Thus to turn Fig. 40 into an interval map the first change necessary is to make the "vertical" lines (warp threads) converge towards the top of the figure as in Fig. 41.

Lengths and intervals

What of the spacing of the woof threads shown in Fig. 40 parallel to Ot and equally spaced along Or?

An observer sitting on the tube OP measures with his own rod the distance between two marks upon the tube in, say, the region $r = CA$ (Fig. 40). He writes

$$ds^2 = dr^2 + r^2 d\theta^2 + dt^2.$$

As he and the tube and his rod are fixed to the rotating disc

$d\theta$ is nought, and as the measurement is a static one (no *time* being necessary in it) dt is also nought and he writes $ds^2 = dt^2$ and $ds = dt$, his rod being a measure of interval. But being in a centrifugal field the rod has contracted, which means that one rod length (i.e. one unit of interval in the Or direction) contains less than one unit of length (in the eyes of the observer O who is drawing the map). This means that into one inch of "horizontal" space in the region A we must put less than one length division—or there will be, in this region, less than one woof thread per inch. Thus the lines parallel to Ot in Fig. 40 should be spaced farther apart towards the top of the figure, as shown in Fig. 42.

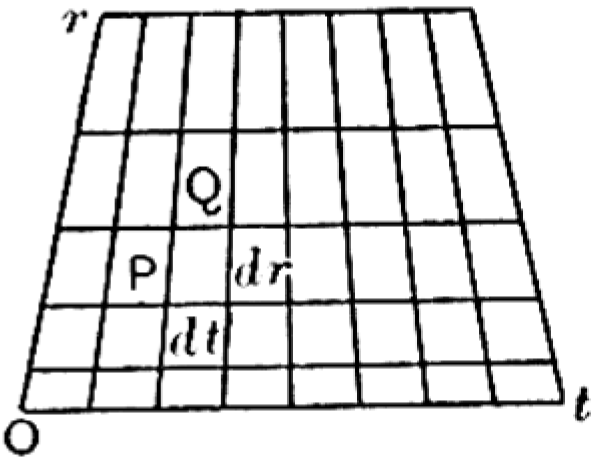

FIG. 42

Fig. 40 is a mere diagram while Fig. 42 is more nearly an "interval map." Such "interval maps" are apt to be found misleading, for although paper distances measured along either the warp or the woof threads are directly proportional to "interval," it is not true that the paper distance between every pair of points on the chart represents the interval between this pair of events. We consider, for instance, two opposite corners PQ of a mesh in Fig. 42. The interval ds between P and Q is given by

$$ds^2 = dr^2 + r^2 d\theta^2 - dt^2$$

and as the diagram refers to the tube OP (Fig. 39) fixed to the rotating disc, $d\theta = 0$ and we have

$$ds = \sqrt{dr^2 - dt^2}$$

and as P and Q are one mesh apart $dr = 1$ and $dt = 0$, whence $ds = 0$. Thus although P and Q are more than one inch apart and the inch is, on this map, the unit of interval, nevertheless the "interval" separating the events P and Q is nil.

It is in fact impossible to make a "true map" of space-time. The best we can do is to draw the mesh-lines in such a way that distances measured with an ordinary scale between pairs of points on any one of these mesh-lines are proportional to the space-time intervals separating the points.

Consideration of such maps makes evident the insuperable difficulty of visualizing the strange geometry of space-time. A striking peculiarity of space-time is this, that, in many cases, the straight line joining any two points represents the longest possible interval path between them; thus, for instance, the interval length of the straight path from A to B (Fig. 43) is longer than any other path, as, say the path ACB. More will be said about this later (p. 250).

Interval maps

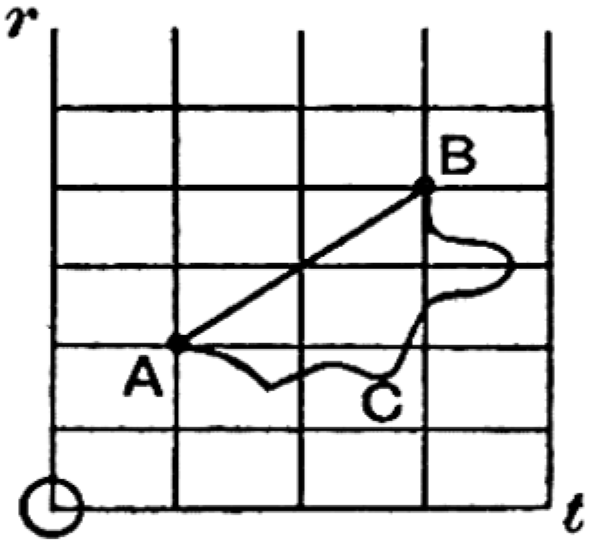

FIG. 43

This and all the peculiarities of space-time as compared with ordinary space result from the presence of the negative sign in the ds^2 equation.

In ordinary plane 2-dimensional *space* we write

$$ds^2 = dx^2 + dy^2,$$

where ds stands for the hypotenuse of the triangle whose sides are dx and dy, and in 3-dimensional space we write

$$ds^2 = dx^2 + dy^2 + dz^2.$$

In *space-time* of two dimensions and of three dimensions respectively the corresponding equations are

$$ds^2 = dx^2 - dt^2$$

and

$$ds^2 = dx^2 + dy^2 - dt^2,$$

while in flat space-time of four dimensions, the equation is

$$ds^2 = dx^2 + dy^2 + dz^2 - dt^2.$$

Always the dt^2 brings in the negative sign which gives its strange characteristics to space-time geometry.

For quantitative treatment it is convenient to work in negative elements of interval area (ds^2) instead of the positive elements defined in the equations just given, and as there seems to be no absolute reason why such an element should be regarded as positive rather than negative the choice of sign is legitimately arbitrary. Thus the ds^2 equation for flat space-time often is written

$$ds^2 = -dx^2 - dy^2 - dz^2 + dt^2.$$

This form has the advantage that when we put our observer at a fixed point in space-time so that, with him, only time is changing, dx, dy and dz being each zero, we can solve the equation, which then becomes $ds^2 = dt^2$, by writing $ds = dt$ and can say that the observer's clock measures *interval*, as thus defined.

$\sqrt{-1}$ If in defining interval the other form of the equation had been taken, in which dt^2 is negative, we should have had

$$ds = \sqrt{-dt^2} = (\sqrt{-1})\,dt$$

and the so-called "imaginary" factor, the square root of minus one, would have spoiled the look of our equations. It can hardly be said that it would have done more than this; otherwise to avoid it in the manner explained would have been illegitimate. There is nothing mysterious about $\sqrt{-1}$. If we choose to multiply all measurable quantities in the world by any numeral or by $\sqrt{-1}$, no inconsistencies will be introduced by the operation. There is nothing "imaginary" about this quantity, though only the imagination can picture its effect when it operates just once upon only one of the members of a group of related quantities. When $\sqrt{-1}$ operates a second time upon a quantity it changes the sign of that quantity. In some branches of applied mathematics, as for instance the vector algebra used in working out alternating current problems, the operation of $\sqrt{-1}$ may be visualized easily; it applies to calculations made by aid of geometrical figures and means "turn the line representing the quantity through a right angle."

Now if elements of interval are the sole "real" elements found in the world it makes no difference to anything measurable if all these elements ds be multiplied by $\sqrt{-1}$.

The invariant property of interval

The same may be expressed by saying that just as the relation $(dx^2 + dy^2 + dz^2 - dt^2)$ between two point-events in space-time is an invariant, that is to say is a relation which gives identical measure results to all observers, so also is the relation $k\,(dx^2 + dy^2 + dz^2 - dt^2)$ where k is *any* constant. It is convenient to set k equal to -1 because then the invariant reads $(-dx^2 - dy^2 - dz^2 + dt^2)$ and becomes, for the ordinary case of two points which are coincident in *space* (so that $dx = dy = dz = 0$), simply $+dt^2$ and $ds = dt$, elements of interval being simply elements of clock-measured time.

"Imaginary" intervals

If, on the other hand, the value $+1$ had been chosen for k in defining ds^2 we should have had

$$ds^2 = dx^2 + dy^2 + dz^2 - dt^2$$

and the interval separating spatially coincident points would be $ds = \sqrt{-1}\,dt$, an "imaginary" interval.

(The idea of the interval *separating* spatially coincident points is repugnant to common-sense because it seems to suggest that there is some sort of separation between a particle and itself. And actually this is the suggestion, for a particle *now* is truly separated from the same particle *later* or *earlier*.)

Even when the value -1 is assigned to k, "imaginary" "intervals" are possible theoretically, for if the observer applies the equation

$$ds^2 = -(dx^2 + dy^2 + dz^2) + dt^2$$

to the measurement of the interval between two points, say, on his own x axis he finds $ds^2 = -dx^2$ and $ds = \sqrt{-1}\,dx$. Practically, however, as will now be shown, such "imaginary" intervals do not occur.

First it should be observed that such an equation as

$$ds^2 = dx^2 + dy^2 + dz^2 - dt^2$$

is meaningless unless some assumption be made in regard to the units in which ds, dx, dy, dz and dt are measured. It is true that $10 = 8 + 6 + 4 - 8$ but it is nonsense to write

10 years = 8 feet + 6 miles per hour + 4 lb. − 8 amperes.

Thus the ds^2 equation has definite meaning, only provided dx^2 and dy^2 and dz^2 and $(-dt^2)$ are all in units of the same sort. But dx, dy and dz are space intervals, measured in, say, kilometre units, while dt is a time interval. To convert *time* into *space* we write

$$\text{space/time} = \text{velocity},$$
$$\text{time} \times \text{velocity} = \text{space}.$$

Hence the assumption involved in the ds^2 equation is that dt is not simply an element of time but

$$dt = \text{an element of time} \times \text{a velocity} = dt \times 1.$$

The assumption then is that dt is multiplied by a velocity which is unity.

We shall see, later, that the velocity by which dt is multiplied in the interval equation is the velocity of light: the equation being written, sometimes, in the form

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2dt^2$$

where "c" stands for the velocity of light. In the simpler form of the equation the units used are such that the velocity of light is *unit velocity*.

In choosing the magnitudes of units in any branch of work the criterion should be convenience. Thus in engineering a unit of force often used is "one pound weight" while in scientific experiments "one dyne" is favoured, although "one pound weight" is nearly half a million times as great as one "dyne."

It is a question of the magnitudes likely to be dealt with. The coal merchant deals in cwts. and tons; the apothecary in grains; the mariner works with nautical miles; the scientist examining the wave-lengths of light deals in microns (each of which is only about one millionth part of a yard); the racing motorist uses miles per hour; the Relativist finds it convenient to use unnamed units of velocity each equal to 186,000 miles per second.

The Relativist's choice of this enormous unit of velocity is justified by the nature of his work. His calculations are concerned mostly with phenomena whose description by the methods of Newtonian mechanics is wrong. Such phenomena are those involving relative velocities comparable with light's velocity (186,000 miles per second = 3×10^{10} cms./sec.). The equation

$$3 \times 10^{10} \text{ cms./sec.} = \text{unit velocity} = \text{unit length/unit time}$$

may be satisfied in many ways. We may, for instance, choose (3×10^{10}) cms. as the unit of length and then one second must be the unit of time; or $(1/3 \times 10^{-10})$ sec. as the unit of time and then 1 cm. must be the unit of length.

The velocity of light called unity for convenience

In considering the *theory* of Relativity the actual choice of units is unimportant, but, in order to make the equations simple by the avoidance of unnecessary symbols, it is well to call light's velocity unity. Then, for instance, the Fitzgerald contraction is written $\sqrt{1 - u^2}$ instead of $\sqrt{1 - u^2/c^2}$. And in space-time diagrams, such as those of Figs. 40 and 42 (to which for convenience we shall refer as *TOX* diagrams), the path of a ray of light will be shown by a line following the diagonals of the meshes. Each side of a mesh is one unit of space or of time, hence in any *TOX* diagram, such, for instance, as Fig. 44, the path *OL* of a ray of light forms the diagonal of the meshes.

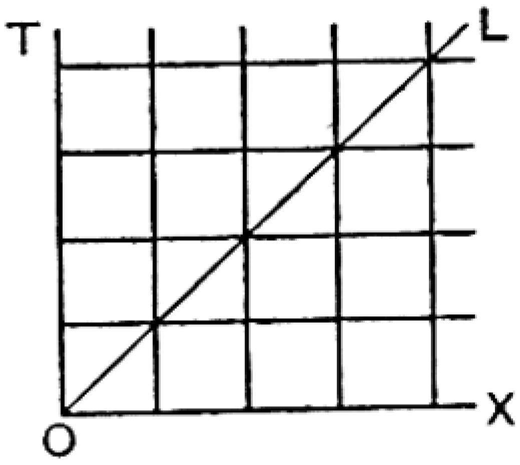

FIG. 44

This condition holds no matter what actual choice of units has been made provided that the *T* side of each mesh represents the same fraction or multiple of one second that the *X* side represents of 3×10^{10} cms.

It is convenient to take equal paper distances for the *T* and *X* sides of the meshes (the units of time and space); then

OL, the path of a light ray, is inclined at 45° to *OX* and *OT* and that paper distance which represents (3×10^{10}) cms. along *OX* must equal that which represents 1 second along *OT*. Sometimes this is expressed by saying that

$$1 \text{ second} = 3 \times 10^{10} \text{ cms.}$$

Bearing these points in mind we may see why it is that by taking the value -1 for k, negative values of ds^2 may usually be avoided.

We refer to the simple case of 2-dimensional space-time for which the ds^2 equation reads

$$ds^2 = -\,dx^2 + dt^2$$

and in which ds^2 can be negative only provided dx is greater than dt, that is to say provided the particle or point-event concerned progresses through an interval whose space component *OM* is greater than its time component *PM*. A particle progressing, for instance, from *O* to *P* in Fig. 45 describes an "imaginary" interval, dx being greater than dt. Such an interval may be called "space-like" because its space component outweighs its time component.

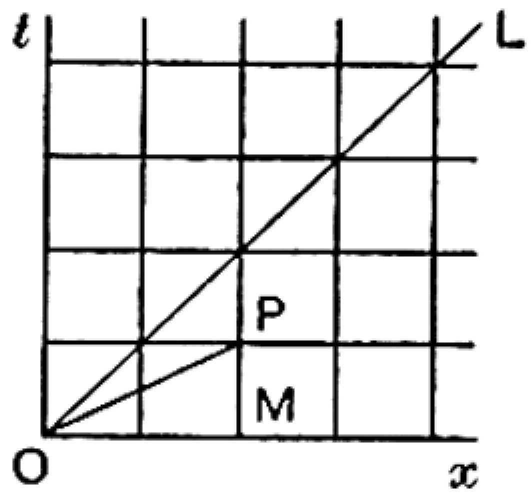

FIG. 45

Inspection of Fig. 45 shows that in order to move along *OP* (through space *OM* in time *PM*) the particle would have to possess a velocity $OM/PM = 2 =$ double the velocity of light. It will be shown later that no massive particle can attain such a velocity (relative to anything identifiable). Similarly in 4-dimensional space-time no particle can move so that $(dx^2 + dy^2 + dz^2)$ exceeds dt^2, and, in general, the use of the negative sign before the factor $(dx^2 + dy^2 + dz^2)$ in the ds^2 equation ensures the avoidance of "imaginary" intervals in most practical problems.

The exception occurs in static problems involving the use by an observer of what is called a "rigid" body.

An observer may define his x axis as, say, a line *OX* drawn upon his laboratory floor. What are we to say about the "interval" separating O from some point P fixed on OX? For both O and P the time t is zero, hence the dt component of the interval separating O and P is nil and we have

$$ds^2 = -dx^2, \quad \therefore \ ds = \sqrt{-1}\, dx$$

an "imaginary" interval.

The occurrence of such exceptions to the rule that by setting the minus sign before the space factor in the ds^2 equation, imaginary intervals may be avoided, results from the remarkable properties attributed by all observers to so-called "rigid" bodies. It is a fundamental claim of Relativity that simultaneity in time at two different points in space is meaningless (because indefinable without ambiguity), but the common notion of a rigid body involves the assumption that all points on such a body are "simultaneous," dt being zero for every pair of them. The weakness in the assumption lies in the fact that the property of "rigidity" is "unreal"; its absolute definition being no more possible than is the definition of shape or of size.

Minkowski's "imaginary" time

The mathematician Minkowski avoided the difficulty of the minus sign by using a quantity $\tau \ (= \sqrt{-1}\, t)$ which he called "imaginary time." Then the ds^2 equation became

$$ds^2 = dx^2 + dy^2 + dz^2 + d\tau^2,$$

a Euclidean equation in four dimensions exactly comparable with the equations used in 3- and in 2-dimensional space.

CHAPTER X

The action of fields of force on light

THE Newtonian ether has been defined as a medium in which radiation travels with the uniform velocity of 3×10^{10} cms. per second. To travel with uniform velocity means to travel in a straight line, for, to be uniform, a velocity must be unaccelerated, and motion, even with constant linear velocity, around a curve, is accelerated motion—the acceleration being v^2/r where v is the linear velocity and r the radius of the curve.

If ever, therefore, it should be found that a ray of light in free "ether" bends, some accelerating force must be looked for that can be called the *cause* of the bending.

Does light bend?

The lift-man will show us that it *must*.

The gauze wall of the lift and the gauze door of the well—past which the wall of the lift slides, in close contact, when the lift falls—shall both be sensitized photographically. Just as the freely-falling lift reaches the well door the lift-man sends a horizontal ribbon of light across the gauze wall of his lift from a lantern fixed some distance from the lift floor on the middle vertical line of the sensitized wall (see Fig. 46). This ribbon of light illumines a narrow line across the gauze wall and acts also upon the door of the well. The lift-man finds—as of course he expected—that the shaft of light has traced a perfectly straight line across the lift wall.

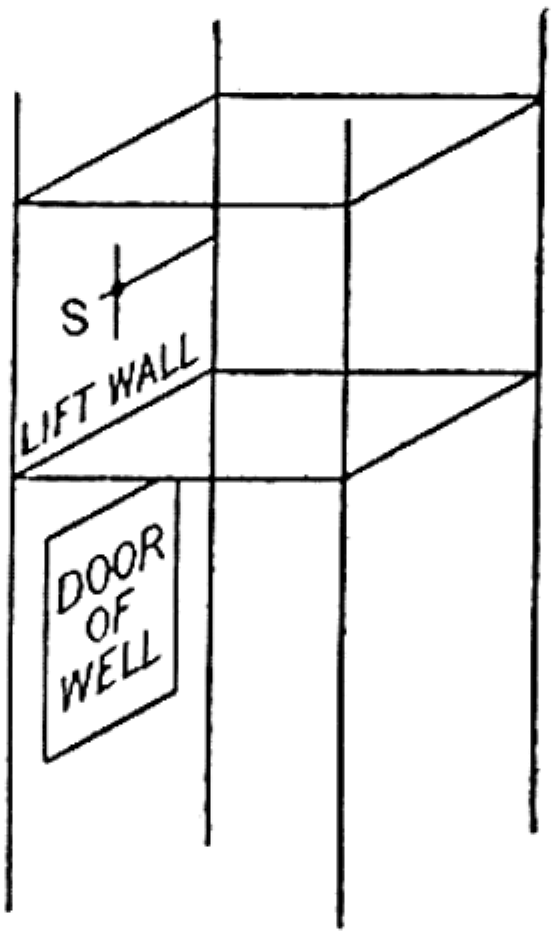

FIG. 46

While the light ribbon, starting at the lantern, travelled across the lift wall, the lift was falling freely, and, precisely in the same way that previously the lift-man's horizontally drawn chalk, tracing through on to the well-wall, described there a parabola, so now must the tip of the light ribbon trace a parabola upon the well door—a downward drooping parabola.

The illustration is an impracticable one because the assumption has to be made that the light at *S* is extinguished at the instant when the light has traversed the wall, so that the whole experiment would last only an infinitesimally short time, but reference to a falling lift does perhaps make the matter more easy to follow.

Thus one and the same beam of light takes a rectilinear path in the lift and a curved path in the well. What can have caused the difference? Not the presence of the lantern in the lift, for it is well known that, after it has left its source, light is propagated independently of that source. Remembering that the essential difference between the lift and the well is that in the former the earth's gravitational field has been eliminated by the motion of the lift, it seems reasonable to conclude that the light ribbon has been attracted by gravitation.

Common-sense says there must be something wrong about such reasoning because gravitational force is real and we are here making of it nothing better than a sort of geometrical fiction. Relativity says that force is indeed a fiction.

Does light respond also to *centrifugal* force?

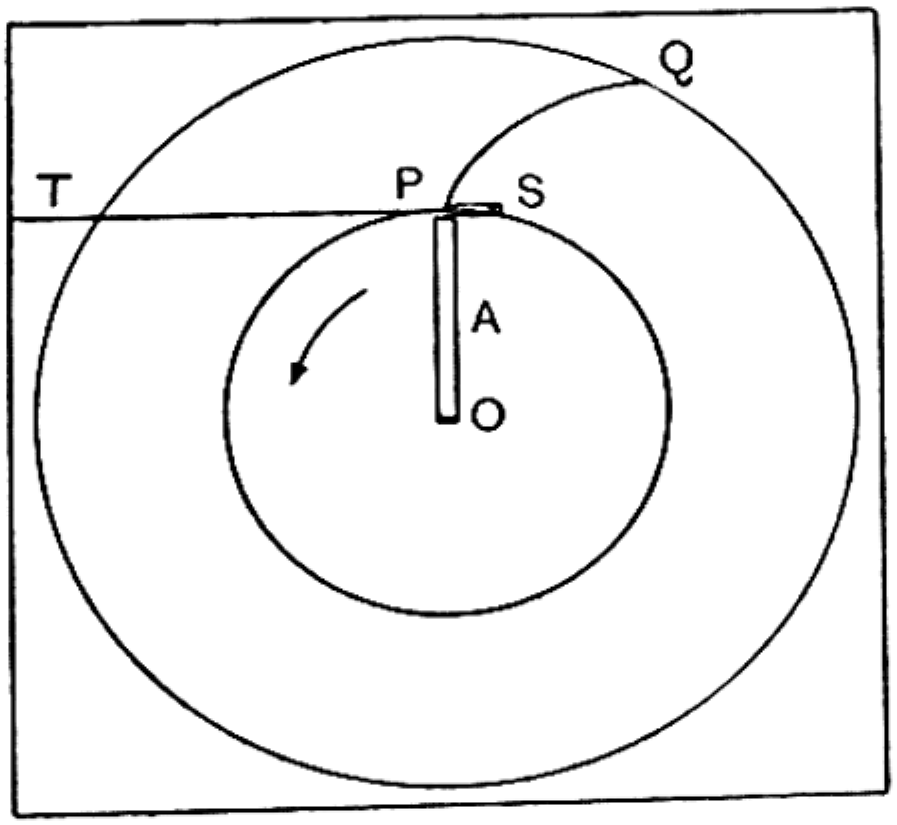

FIG. 47

We suppose that a rotating arm *A* (Fig. 47), carrying a flash-light source *S* sending a beam along a little tube *SP* perpendicular to *OP*, rotates with uniform angular velocity about the centre *O*. The table *T* beneath the rod is fixed, and at the instant when the arm *A* comes into the position shown in Fig. 47 a momentary flash of light is emitted by *S* at *P* from the tube *SP* (which is fixed rigidly to the arm *A*).

In respect to the table *T* the tip of the light ribbon sent out from *P* will describe the straight line *PT* tangential to the circle described by *P*, but in respect to the disc (which we may imagine is carried by the arm *A* to provide a background) the tip of the light ribbon must describe the outward sweeping curve *PQ*.

Thus one and the same beam of light travels rectilinearly on the table *T*—where no centrifugal force exists—but in an outward sweeping curve on the rotating disc, on which centrifugal force exists.

It seems reasonable to conclude that light is thus affected by centrifugal force. Moreover, since the reasoning by which the shape of the path *PQ* would be deduced for the tip of the light ribbon is precisely such as would be used in deducing the path of a particle similarly shot off (with light's velocity) from *P*, we may conclude that the action of centrifugal force on light is the same as its action upon material particles.

Reasoning of this sort is legitimate but we are not warranted in concluding from it that the amount of the deflection of light in gravitational fields (of which centrifugal fields are a special case) may be calculated in accordance with Newton's laws. The conclusion is indeed warranted that the path of a ray of light and that of a particle of matter having light's velocity are similarly affected by fields of force, but recent experiments have shown that they both are affected to twice the extent which Newton's laws predict.

Astronomical observations made in May 1919 have shown conclusively that light is deflected by *about* twice the amount by which, according to Newtonian mechanics, a particle having light's velocity should be deflected, and Einstein's law of gravitation predicts that the deflection must be exactly twice that amount.

Assuming now that we are justified in stating that light is deflected in a gravitational field by twice the amount which Newton's laws predict, we proceed to draw some striking conclusions by reference again to the experiment of Fig. 46.

First we shall show that the assumption leads to an obvious "absurdity," repugnant to common-sense; then that common-sense had made an unwarranted assumption, and finally we shall deduce interesting conclusions about the behaviour of rods and clocks in gravitational fields.

A source of light *S* is fixed on the gauze wall of the lift so that it can send out a horizontal ribbon of light from *P* (a point on the vertical centre line of the lift). The "ribbon" of light is seen on edge as the line *PQ* in Fig. 48. This ribbon by its breadth penetrates the gauze and so strikes simultaneously the sensitized door of the well down which the lift falls. The word "simultaneously" introduces here no ambiguity because the lift wall and the well door which are to be affected simultaneously are in close contact while the lift moves down the well.

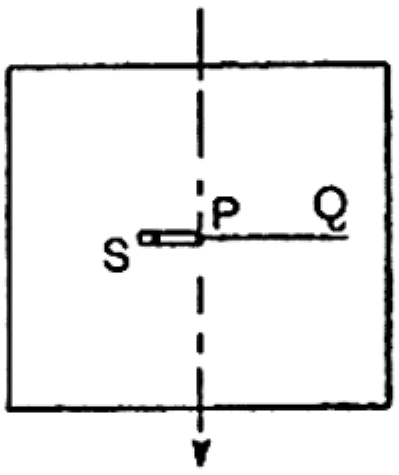

FIG. 48

Since the lift is falling freely there is no gravitational field within it; there is thus no field of force in the frame of reference constituted by the gauze wall of the lift and the tip of the ribbon of light must describe a true straight line across the lift wall.

It is known that the tip of a ribbon of light takes time to travel—just as a bullet does. Thus on turning on the light at *P* the tip of the ribbon of light begins at *P* and travels like a bullet, with uniform velocity 3×10^{10} cms./sec. from *P* to *Q*. Always the lift and the lantern and the straight ribbon of light travel together *en bloc*, and the tip of the ribbon marks its trace upon the sensitized door of the well. If we know precisely the speed of the lift down the well at each instant we can describe exactly the path traced upon the well-wall. The argument determining this path is exactly such as would be used for a bullet coming with light's velocity from *P*.

Relativity says the path on the well-wall will dip down twice as far as that path does, which the above argument tells us the light *must* trace upon the well-wall.

We go step by step once more through the argument and find that, in order to reconcile it with the Relativist's contention, we have to suppose that the tip of the ribbon descends twice as far down the well as the lift itself does, *that is to say the ribbon droops downwards on the lift wall instead of being "straight."*

Such a supposition is inconsistent with our initial supposition that the path of the ribbon is a straight line parallel to the lift floor.

Thus the ribbon is at the same time both "straight" and "bent."

If the ribbon did not bend it must we know trace on the well-wall the Newtonian parabola, or, conversely, a bullet following the Newtonian parabola with light's velocity on the well-wall would trace a horizontal *straight* line on the lift wall. We know exactly what this means and can, indeed, with certainty, draw beforehand in the stationary lift the *straight* line which the Newtonian bullet must trace on the lift wall. For even if we allow that the Newtonian law of gravitation may be inaccurate we can claim confidently that the error must be far less than would produce the required difference of 100 % between the predicted and the actual distance fallen by the lift.

Relativity says the light ribbon (or a bullet having light's velocity) will not trace the Newtonian parabola on the well-wall but will trace a curve having double as much droop. We conclude, therefore, that neither will it trace a *straight* line on the lift. Yet by assumption a beam of light in free space with no field of force (as in the falling lift) *must be straight.*

Ambiguity of the word "straight"

Does not such inconsistency prove Relativity to be wrong? It does so, provided we are using our words consistently. But what if the word "straight" should be ambiguous?

The Relativist explains the lift difficulty in the following way: The *straight* line which you draw in the stationary lift, with full confidence that it is straight and also that it must represent the Newtonian bullet's path in the lift when this is falling freely, is truly straight in the sense that it is the shortest path between its beginning and its end while the lift is stationary; but when the lift has been freed from gravitation the line you have drawn will appear *to you* (as you look into the falling lift from *terra firma*) as no longer "straight," although it will be straight as seen by the lift-man. A ray of light sent from one end of the line to the other, or a bullet having light's velocity would be found to trace this path which now you call bent.

To make the matter more clear let us begin with the ribbon of light in the freely-falling lift. As a Relativist I claim that the path traced on the well-wall by the tip of this ribbon as it passes from *P* to *Q* (Fig. 49) is not the Newtonian parabola but is a much deeper drooping curve. In fact (we having entered the freely-falling lift) I will draw for you the line which the tip of the ribbon would have to follow in order to trace the Newtonian parabola on the well (he draws *PQ'*). To us in the falling lift *PQ'* is not straight but to the ground-man looking into the lift it is straight while *PQ* droops downward.

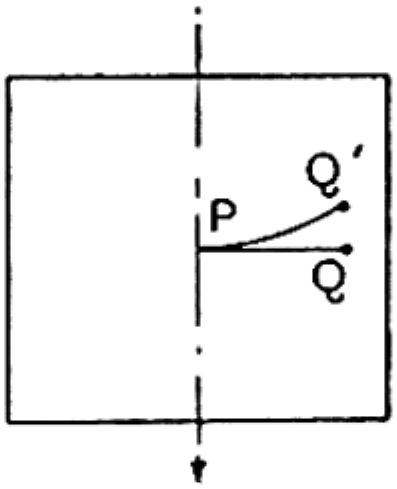

FIG. 49

To use the language of people who believe in the Newtonian ether, we may say that the lift suffers distortion in a gravitational field, actually contracting along the lines of force.

To a distant observer looking on from outside earth's field, it is the stationary lift which has suffered contraction relatively to the undistorted freely-falling lift, but owing to the peculiar reciprocity of all the changes of dimensions which, from the common-sense point of view, appear to be demanded by Relativity, the ground-man looking into the falling lift attributes *contraction* to it, in the same way that the lift-man looking out from his falling lift attributes contraction to the things he sees on *terra firma*.

Contraction of bodies in a gravitational field

It is important to note that the contraction is along these lines of force—which are not parallel but converge towards the earth's centre. Uniform longitudinal contraction of, say, this page, would not cause the lines of print to bend, but if the paper had a converging grain and the threads of the grain all contracted in length by an equal percentage then a bending of the lines of printing might ensue, the lines becoming concave towards the centre of convergence.

This is an interesting point, and by following the line of thought which it raises one can realize that, on a flat earth—on which all the lines of force of the gravitational field would be truly parallel—light and high-speed particles would be affected by gravity just as slow-speed particles are, namely, in accordance with the Newtonian law.

The intensity of such a field (since the lines would be parallel) would be the same at all distances from the flat earth. There would thus be no hummock in space-time; gravitation would be unreal, it could be explained away everywhere by taking a uniformly accelerated frame of reference and there would be no point in saying it "existed" at all.

Your argument by which you showed it to be absurd that light should follow a steeper curve than the Newtonian parabola was due to your neglect of the fact that the space in a gravitational field is non-Euclidean. Of course all the phenomena, those of slow particles, which are satisfactorily described by Newton's laws, and those of high-speed particles, whose description demands the more inclusive law of Einstein, are effects of space-time distortion, but this complex distortion involves a pure space distortion (detectable theoretically in static experiments by the non-Euclidean results obtained, say, by examining the geometry of a circle lying in a gravitational field) which is not taken into account in Newton's laws, and it is this space distortion which expresses the chief difference between Einstein's and Newton's law of gravitation.

Einstein's law introduces no such idea as the attraction of matter by matter. Particles and rays have paths in space-time.

Nature has laid down these paths and seems to have followed certain rules in regard to them, or to have adopted a certain definite sort of ordering of these paths in space-time. Einstein's law of gravitation is intended to describe this ordering.

According to Relativity, but using the language of the "ether" convention, rods become shortened when laid along the direction of the lines of force of a gravitational field. In order that, in spite of this change in his measuring rods, an observer shall still obtain correct measures of space-time intervals it is to be expected that a gravitational field must have a complementary effect upon the going of clocks.

In considering this point we start again with the assumption that light is affected by gravitation. When traversing a gravitational field (across the lines of force) light is bent and this bending may be explained (see p. 256) as a refraction effect due to a reduction in light's velocity—much as in the case of ordinary optical refraction. It is to be understood, however, that the reduction in velocity is appreciated as such only by a distant observer looking on from outside the gravitational field, and is not inconsistent with *the principle of the constancy of the velocity of light* as actually measured by an observer located in the region which the light is traversing. An observer on the sun would find for C the same value 3×10^{10} cms./sec. which is found locally everywhere. Theoretically the sun man might be aware of a slight inconstancy of the velocity at different levels above the sun's surface, because of the falling off of the intensity of gravitation, the sun man being able to look as an external observer into these higher regions, but the detection of the effect of the diminution in gravitational intensity could only be by indirect measurement. The Principle of Equivalence demands that an observer inhabiting a gravitational field must obtain the same value for the measured velocity of light as another situated far from gravitating matter. The velocity of light, directly measured, is always and everywhere unity.

The Principle of Equivalence

The Principle of Equivalence says that inside an accelerated lift located in a region free from gravitational fields due to matter, there exists the precise equivalent (over a sufficiently small floor area in the centre of the lift) of a gravitational field. Tacitly we have assumed the truth of this principle when claiming that, in a freely falling lift, the earth's field has no effect, for we have then assumed that the downward acceleration of the lift creates a field of force precisely equal and opposite to the earth's gravitational field.

We imagine that there are two lift-men *A* and *B*, each in his lift and both located in a region free from gravitation. *A* is "at rest" and *B* has *just begun* to move away from *A* with an acceleration equal to that of the sun's gravitation as experienced by the sun man. The point in saying *B* has "just begun" to move is to enable us to assume that *B* has an acceleration equivalent to a gravitational field, yet possesses no appreciable velocity relative to *A*. We thus avoid having to allow for a FitzGerald contraction of rods and slowing of clocks which would combine themselves with the gravitational contraction and slowing we now are discussing.

A and *B* are equipped with properly standardized rods and clocks. If *A* measures the velocity of light he finds it to be 3×10^{10} cms./sec. for he is "at rest" and is disturbed by no gravitational or equivalent field. If *B* measures the velocity of light it seems possible that he may find some other value because, for aught we know, the velocity of light may be affected by *B*'s acceleration since this is the equivalent of a gravitational field. Suppose we suggest that *B* may find a value C' instead of C; what will *B* say to our suggestion? He will tell us that our ideas are confused. He (*B*) *knows* that he himself is at rest, having confirmed the fact by measuring light's velocity and finding it to be C. It is *A* who is suffering acceleration, and the suggestion about finding the value C' must be referred to *A*.

The point is that in affirming *A* to be at rest and *B* suffering acceleration we are making an arbitrary choice. The acceleration is relative and *B* is equally as well warranted as *A* in regarding himself as "at rest." Such indeed is the fundamental standpoint of Relativity. If, as we suggested, A and B should find different values for C the conclusion would be legitimate that A and B were in different states of "absolute motion," or of motion relative to the Newtonian ether.

By such reasoning, therefore, we have to conclude that both A and B must find one and the same value for C, and since, by the Principle of Equivalence, either A or B may be regarded as in a local gravitational field while the other is in a region free from gravity, we conclude that a gravitational field can have no effect upon the velocity of light *as measured locally* within that field. It is easily seen also that the velocity of light must be the same in all directions in either lift, and that, therefore, in saying that the velocity of light as measured locally must be constant we mean that it must be the same for a ray travelling across the lines of gravitational force as along them.

In order now to reconcile these conclusions with the Relativist's deduction that light rays are bent and that, therefore, light's velocity suffers diminution in a gravitational field we are led to conclude that, in such a field, *clocks must go slowly*.

For we consider a ray travelling *across* the sun. The sun man sets about measuring its velocity. He holds his rod lengthwise along the ray and notes the time it takes to traverse the rod. We have agreed that he must find (length/time) equal to 3×10^{10} cms./sec. Yet we know that, held transversely, in a gravitational field the rod is one unit long (since rods are affected by gravitational fields only when oriented along the lines of force), and the Relativist tells us that *as seen from afar* the ray's velocity is less than 3×10^{10} cms./sec. The only conclusion possible is that the sun man's clock indicates fewer "seconds" than the light really took to traverse the rod, that is to say the clock goes slowly.

The reason why a transverse ray of light has been chosen rather than a ray travelling radially relatively to the sun is that there exists no direct experimental knowledge as to the behaviour of a radial ray, while that of transverse rays is known with fair certainty. It is interesting to consider, however, the information as to the behaviour which a radial ray must exhibit, derivable from the acceptance of the conclusions already reached as to rods and clocks. In order to measure light's radial velocity the sun man lays his rod along a radius of the sun—and therefore along a line of force—and uses, perforce, his slowly going clock. Adhering to our previous conclusion that, no matter what happens to rod and clock, the sun man must find the value 3×10^{10}, we conclude now that as the rod has contracted (so saying the ray goes farther than actually it does in traversing the length of the rod) and the clock is going slowly (so saying the ray has taken fewer seconds than actually it does) the sun man *must* find a value greater than 3×10^{10} unless in fact *light goes with ever increasing slowness as it approaches the sun*. Thus we infer that a radial ray suffers repulsion as also would a high-speed bullet going towards the sun radially.

Is the gravitational contraction real?

We. You speak of the contraction of bodies in a gravitational field. Is this contraction real?

Relativist. It is as real as any change of linear dimensions can be, but from the universal observer's point of view it is not real. How can a change of length be real when length itself is relative? The only real thing in the gravitational field is the hummock in space-time, of which the properties of the field (its geometry) are the expression.

The hummock cannot be described strictly in terms of things which it is not, such as changes of length and slow going of clocks, and in attempting thus to describe it our reasoning has been inaccurate. We have spoken, for instance, of a change in the dimensions of the lift, and the implication seems to be that this change in dimensions must involve stresses in the material of the lift—stresses which should produce a slight buckling or warping of the material. Yet no such implication is intended. A buckling such as one has in mind is one which may bring into contact pairs of particles in the lift structure which previously were not in contact—

that is to say it may produce new space-time coincidences, new "realities." But such a notion is incompatible with Relativity. Gravitational force is a relative thing; I can specify an observer (namely the lift-man in a freely-falling lift) who will be unable to observe the force, and another near him, namely the ground-man, who does observe it; yet they both do observe the space-time coincidences (contacts of material particles) which define the thing called the lift. These coincidences, being real, are similarly observed by both observers. (By considering the instant at which the lift is just passing the ground-man—and so is at rest relatively to him, the rate of change of its distance from him being at that instant nil—we have the condition that the lift is then at rest relative to both observers. Thus no question of relative motion, involving FitzGerald contractions, arises.)

Thus if one man is observing a warped lift and the other an unwarped one, the warping must be a pure distortion (cf. an image in a distorting mirror) without creasing—that is to say it must be a distortion which, while producing arbitrary percentage changes in the distances between any or every given pair of particles, can never reduce any of these distances to nought (so as to produce a new contact). And if the idea of stresses in a material involves that of new contacts, the "contractions" produced by gravitation cannot be such as to produce stresses.

The difficulty involved in the attempt to explain the hummock that underlies gravitation, in terms of linear changes, becomes evident by such a consideration as the following: I take a bicycle wheel and put a massive body near its hub. The gravitational field due to this body is roughly along the spokes; hence the spokes contract. The rim, lying across the lines of force, suffers no contraction—no change in circumference. Surely then the spokes must become stretched—and having stretched must stretch again, and so on until they break. Such a notion is incompatible with Relativity. It is better to avoid talking about the contraction of the spokes.

What "really" happens is a change in the measure relations (the geometry) of the region defined by the space-time coincidences forming the wheel.

If, however, you will drag me into refined difficulties let us return to the lantern in the falling lift. The first inaccuracy was that of saying that the path on the lift wall, corresponding to the Newtonian parabolic bullet path on the well-wall, could be drawn, beforehand, as a *straight* line in the stationary lift. Strictly, Newtonian mechanics would demand an arc—very nearly an arc of a great circle of the earth. This is because the bullet's acceleration is radial and not parallel to one given radius of the earth. Thus if you drew what you considered to be a true straight line in the stationary lift, you would find that the locus of a bullet, which the groundman fired horizontally across the lift at the instant when the lift began to fall, would not be the straight line you had drawn, but that this locus would curve down below the straight line a little. *OM* in Fig. 50 would be the line you had drawn, and *OM′* the locus of the bullet. According to you, however, *OM* would certainly prove to be the path of the *light* ribbon spoken of on p. 189.

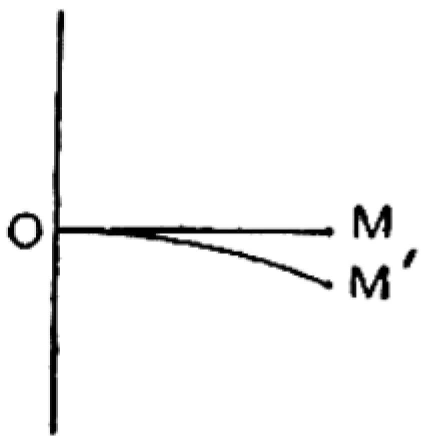

FIG. 50

According to this point of view Newtonian mechanics demands that light shall be affected by gravitation nearly, but not quite as much as a bullet.

Usually it is said that Newtonian mechanics recognizes *no* effect of gravitation upon light. It seems impossible, however, to maintain such a view without ambiguity. If light travels in a straight line in space where gravitation exists, as well as in space free from gravitation, how are we to describe the behaviour of the ribbon of light? We imagine the ribbon split longitudinally by the gauze into two strips, one of which finds itself outside the lift in a field of gravitation while the other strip finds itself inside, where no gravitation exists. To maintain that the light travels in a straight line both inside and outside the lift is to maintain that the tip of the one strip of the ribbon parts company with the tip of the other strip, the one tip describing a straight line on the lift-wall while the other describes a straight line on the well-wall. The absurdity of this is evident when we reflect that, having left its source, light travels independently of that source, and that, therefore, our ribbon of light knows nothing of what may be happening to the lift, and certainly cannot know that it ought to split itself in two when the lift happens to fall.

We must admit, therefore, that Newtonian mechanics expects light to be affected by gravitation, at any rate almost as much as it expects a bullet of equal velocity to be affected.

Another difficulty: You are not quite happy about the converging grain of the sheet of paper. Contraction of the grain might bend the lines of print, but whether upwards or downwards depends on what section of, or line in, the page we regard as fixed. Perhaps Fig. 51 may help us. *E* is the earth and *S* is the gauze wall of the lift forming in imagination a piece of a great equatorial disc of gauze. The "keystone"-shaped outline of the gauze wall of the lift is thickened in the figure. The strands of the gauze are radial to earth in the warp and concentric with earth in the woof.

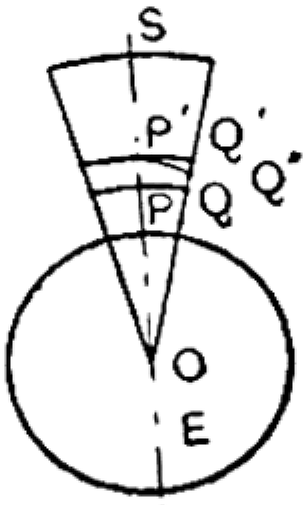

FIG. 51

With gravitation present all the radial wires have suffered contraction, but when the lift begins to fall the contraction is relieved. Thus the points *P* and *Q* on the circle, which you drew in the stationary lift and predicted would be the bullet's locus in the falling lift, rise to *P'Q'*, because the radii *PO* and *QO* now lose their contraction. Clearly the curvature of *P'Q'* is less than that of *PQ*.

Owing now to the reciprocity of such effects (as discussed on p. 194), when the ground-man looks into the falling lift he sees contraction to have occurred, and to him *P'Q'* is below and more bent than *PQ*, that is to say the bullet's path droops further than he thinks it "ought." He finds that the bullet has not obeyed Newton's law but has suffered a greater deflection. How much greater depends upon the "horizontal" velocity of the bullet and when this equals light's velocity the curve *PQ* will have twice as much droop as *P′Q′* has.

Yet this argument also is loose and inaccurate; for, as we have said already, we must not allow spokes and circumference to contract and expand together, otherwise the region remains Euclidean.

The inadequacy of explanations of Relativity in ordinary language

What has been said is only an imperfect suggestion of the reality—a suggestive evasion of the facts. These facts are in a space-time world, and no language except that of mathematics can talk logically about such a world. The distortion in space-time which underlies a gravitational field cannot be pictured as a change of curvature of a line or of a surface; it is something much more complex.

CHAPTER XI

The Special (or restricted) Theory of Relativity

THE Special—sometimes called *restricted*—Theory of Relativity is that part of the General Theory which applies to forces which can be explained away everywhere simultaneously without the postulation of a hummock in space-time: it deals with phenomena whose description is possible in the language of Euclidean geometry.

In Newtonian terms the Special Theory is, that it is impossible by any conceivable experiment to detect uniform motion through the ether.

So much has been written about this special part of the General Theory of Relativity, that it is difficult to say anything new. The endeavour will be to deal with the matter from a point of view which makes rather clearer than usual the oneness of the Special Theory with the General.

The basis of our introduction to the General Theory was the assumption of the non-existence of an identifiable medium for the transmission of light; an assumption which was seen to be implied necessarily in the more vague assumption that *absolute space* is non-existent.

Matter is thought of as needing space in which to exist, and the amount of the space it needs is found by measuring the matter which occupies it. Thus space and matter, separately, seem to be indefinable. There was, however, one thing we knew about space apart from its occupation by matter, namely, that it formed the seat and also the medium of transmission of radiation.

If the position and dimensions of the measurable matter which occupied space defined for us the position and dimensions of the space occupied, it was to be expected that the measurable properties of the radiation (e.g. light) transmitted by space would tell us something else about space. Unfortunately the properties of light proved "absurdly" inconsistent with the notions we already had formed about space in terms of the properties of the matter which occupied space. The phenomena of light transmission are inconsistent with the attribution of either dimensions or motion to space. Light travels with the same velocity relative to all observers, no matter what may be these observers' motions relatively to one another; hence all observers must be travelling at the same velocity relatively to the medium which transmits light; hence this medium cannot be unique but must be a thing without identifiable parts.

Assumption of the constancy of the velocity of light

We take now as basis for the consideration of the Special Theory the assumption that scientists are justified in interpreting phenomena—notable among which is that dealt with in the Michelson-Morley experiment—to mean that, with extreme precision, the velocity is the same relative to all frames of reference, no matter with how great (uniform) velocities these frames may be moving relatively to one another.

On first meeting this assumption we find it so "obviously absurd" that, rather than make it, we prefer to assume that all the experiments which seem to demand it must be wrong; for, however broadminded we may try to be, there are limits to our credulity. We refuse to believe, for instance, that a thing can both be and not be at the same time. Even Bertrand Russell who, in common with all mathematicians, objects to knowing what it is that he is talking about or whether what he is saying is "true," insists on postulating this. Yet what is said about light robs us of our right even to this postulate. It is said, for instance, in effect, that if you start out to travel, in contact with the tip of a beam of light, actually sitting upon its wave front as it spreads out through space, and are foolish enough then to set about measuring with clock and scale what is the velocity relative to yourself of this thing upon which you are sitting, you will obtain the result 3×10^{10} cms./sec. Now, surely, to be sitting upon the wave front means to be at rest relative to it, while to have a velocity means to be not at rest. Hence you are being asked to believe that you are at the same time both at rest and not at rest relatively to one and the same thing.

Examination discloses, however, that an unwarranted tacit assumption is at the root of the difficulty in accepting as a fact this curious property of light. The assumption is that the measuring rod and the clock which you carry on your journey retain unchanged their properties of length and of time keeping. What if this very high speed makes your clock go with infinite slowness? You are sitting on the front of the train of light waves, and you look back (do not ask how!) over the long train. Your slow going clock tells you those waves behind you actually are approaching at 3×10^{10} cms. per second.

The Relativist is not surprised, however, if you find the position unsatisfactory; he considers you have made a still more fundamentally unwarrantable assumption, namely, the assumption that you "started out" to travel with the light wave; for he considers that such "starting out" would have been found impracticable. He is prepared to show that the inertia of your body would be infinitely great if you were moving with light's velocity so that no "force" would have sufficed to impart that velocity to you. To make practicable such a journey as you have in mind you would have to travel at least a little slower than the light waves. The effect upon clock and scale of your velocity through the ether would be such that when now you set about finding how long a time the light overtaking you took to traverse your measuring rod you would surely obtain such a value for the time interval as would make the velocity of light relative to yourself 3×10^{10} cms./sec.

The universality of the value of the velocity of light relatively to the observer who measures it (and light can have no other *real* velocity, for a *real* thing means, in science, a *measured* thing) is an experimental fact, but to call it such is meaningless unless we assume that all possible observers carry "corresponding" scales and clocks—that is to say scales and clocks which agree with our standards when they and the standards are brought to rest and placed so close together that they may be regarded as coincident.

Measuring rods and clocks

A measuring rod seems hardly to need definition, but a clock is a complex thing. It would be allowable for the Relativist to say: I do not know what you experimental physicists mean by a clock, but so long as I may safely assume that your deduction of the constancy of light's velocity is correct, I have no interest in the construction of your clock, for I have a much simpler clock of my own. It consists of a rod along which light takes unit time to travel. You have assured me that, no matter how quickly my rod may be moving relatively to the source of light, or in which direction along the light rays, the light will take exactly one unit of time to traverse the rod. My rod is thus a clock. You say it will be a little unwieldy, since its length will have to be 3×10^{10} cms. (= 186,000 miles), but such merely experimental difficulties do not concern me. I do not wish to make experiments; you can do this with your own clocks, since they are less unwieldy than mine. My object is only to deduce, without further experiment, the consequences which seem bound to follow consistently with your experiments, and to make predictions whose accuracy shall be probable.

If you will insist that at least I may explain a conceivable way of making my unwieldy clock practicable, I ask you to give me a rod one yard long, with a mirror at each end, the two mirrors to be parallel, perpendicular to the rod, and facing each other. A bright spot of light in the middle of one mirror sends out a ray which, striking the opposite mirror, is reflected back along its own path, then forward again and back once more and so on, like the bob of a pendulum, to and fro for ever. The "beats" of this pendulum will be exactly that fraction of a second which one yard is of 186,000 miles. I refuse to worry about how the beats are to be recorded or counted.

The rod-clock

The beats are there, and you assure me they always must represent the same numerical fraction of a unit of time no matter how my "clock" may be moving. I can now measure time accurately, and *can express time units in terms of units of length*. When the light has made 1760 × 186,000 journeys from mirror to mirror, one second will have passed, so I call this number of yards one second.

The physicist can see no fault in the Relativist's reasoning, but is worried by the fact that, in explaining the Michelson-Morley experiment, which provided one of the most direct proofs of the constancy of light's velocity, it had seemed necessary to make two assumptions, 1st, that a measuring rod contracts in the direction of its motion (but not transversely), 2nd, that a moving clock goes more slowly than one at rest.

If the moving Relativist uses his rod as a clock while orienting it in the direction of motion, then, since his rod has become shortened, light ought to traverse it in less time and so the Relativist's second would be too short.

He realizes that the difficulty is a "common-sense" one, and that the way to obtain common-sense explanations is to use the ether analogy or its equivalent (such, for instance, as that of a train on an embankment); moreover even the Relativist grants that, when used with care, the ether analogy leads to correct conclusions. The ether explanation of the FitzGerald contraction and the consequent slowing of clocks is given in some of the text-books on Relativity, but nowhere is it given more clearly and concisely than in Professor Eddington's "Report on the Relativity Theory of Gravitation" (Physical Society, 1918; Fleetway Press). It will there be seen that the assumption of the FitzGerald contraction involves the further assumption that moving clocks go slowly (a slowing which is quite independent of that suffered by clocks when stationary in a gravitational field) and that, in addition, the *setting* of a distant *retreating* clock (retreating relatively to the observer) is *behind* that of a clock, similarly moving, nearer to the observer, while a distant approaching clock is *ahead* of a nearer one.

These relative settings are no less important than are the slow going and the contraction which demand them, and it is the failure to take account of clock-setting or its equivalent which causes the apparent difficulties in such arguments as that by which the rod-clock appears to be wrong.

Careful re-consideration of these points convinces the physicist that it is indeed legitimate for the Relativist to use a rod as a clock in the manner he proposes. It would be more ideal to make it of infinitesimal length *dl* instead of one yard but, as a rule, space is nearly "flat" over an area so small as one square yard. The point is that the length of the rod ought to be so short that the rate of going of an ideal clock placed at either end of the rod should show no "gravitational" difference. Orientation can have no more effect upon the readings of the rod-clock than on those of an ordinary stop-watch, for the rod-clock, like any other clock, may be said to "go slowly" when it moves and its setting will adjust itself in accordance with the ether argument. If the idea of a "rod-clock"—a mere light pendulum—having a setting, gives difficulty, it is only necessary to suppose that the light, on its journeys to and fro, actuates, somehow, a recording device having scale and hands like those of an ordinary clock.

Better, however, than trying to picture, by aid of the ether analogy, the way in which the apparent shortening which an observer *S* sees in the rod of a moving observer *S′*, and the apparent changes in *S′* clock, are consistent with the fact that both *S* and *S′* find one and the same measured value for light's velocity, is to examine the *TOX* diagram of Fig. 57 which is explained on p. 218. The ether and its direction of flow are artificial conceptions which have to be discarded later in the argument; they are replaced in the *TOX* diagram by the space-time continuum and the orientation within it of the *T′OX′* frames of variously moving observers.

The *TOX* diagram explains at once the consistency of the Fitzgerald contraction (suffered by *S′* as seen and *accurately measured* by *S*) with the fact that *S′* does indeed use a *greater* unit of length than *S* does. Both the contraction of *S′* rod and the expansion of *S′* unit of length are *as seen by S*. The apparent inconsistency of saying, as *S* does, that *S′* rod contracts and at the same time *S′* is using a unit of length which has expanded, may be explained away easily. When S measures S' rod he does so in his own (S) units of length; but when S speaks of the unit of length which S' uses, S expresses the value of the length of that unit by means of a numeral representing a quantity which to his (S) mind contains some time as well as space. The sloping space sides (parallel to OX') of S' meshes represent pure space intervals to S', but to S they represent a combination of space and of time.

S' can say, and truly, that his own rod is of true standard length and represents the true unit of length. Looking at S he finds (as S does for him) that S uses a contracted rod but an expanded unit of length.

Again, when either (say, S) watches the other (S') using his rod, he finds that the ends of that rod (since, relative to S, it is a *moving* rod) are separated from each other *in time as well as in space*, although S' thinks they are separated in space only. S adds together, according to the geometry of his own (S) space-time, the time and space components which appear to him (S) to separate the ends of S' rod, and, finding a result numerically greater than his own (S) unit of length, S says S' is using an expanded unit of length.

In regard, however, to *S' rod*, S says that as the two ends of S' rod are separated in time as well as in space (S time and S space) I must separate out the space in order to say what has happened to the *length of the rod*. To separate out the space means to project S' rod (represented by one of the space sides of S' meshes) upon S space axis (OX). Reference to the diagram shows that this projection is less than one of the space sides of an S mesh; the defect representing the Fitzgerald contraction.

Put briefly the ground of the difficulty underlying the description of the behaviour of moving rods and clocks is this, that observers in relative motion attach different meanings to the terms "space" and "time"; that which appears to the one as a pure space interval appearing to the other a mixture of space and time. The only two things upon which all observers agree are the numerical values of the velocity of light and the magnitudes of "intervals."

Reasons for expecting the velocity of light to be constant

The constancy of light's velocity is an experimental fact and the Relativist has a right to use it as such, but it is very interesting to consider what the Relativist would do if he knew nothing, from direct experiment, of the constancy of light's velocity (which, for convenience we may call "*C*") and took as a basis for his reasoning simply the assumption that the constancy of the dimensions of bodies and of the beats of pendulums was merely a convenient arbitrary assumption and undemonstrable; would he not be led to some conclusion in regard to light's velocity?

He would have to assume that, since light possessed a velocity, this velocity either would be different for different observers or the same for different observers. The possibility of "*C*" being inexplicably different at different times to one and the same observer (that is to say to an observer—himself of negligible gravitational mass and so unable to affect light's velocity by his mere presence—moving always in the same manner relatively to the same light source) need hardly be entertained, for if it were so optical science, and also, by inference, electromagnetic science—and indeed, probably, science of whatever sort—would be impossible: there would be nothing to observe but a confusion of events.

Now if *C* were to be different to differently moving observers there would be nothing by which to account for the difference except the different motions of the observers, and since it is known that not the source of light affects a distant observer but rather the waves that reach him from that source, it would be the difference in the motions of the observers, relatively to the waves of light, that would have to be used in describing the differences of *C*. But the waves are located in the "ether," hence the differences in *C* would have to be interpreted in terms of the differences of the velocities of different observers relative to the ether.

To be able to measure a velocity relative to a thing is to be able to identify it, to the extent, at any rate, of being able to use it as a frame of reference, and it is meaningless to speak of a velocity in reference to a frame which, like the Newtonian ether, has no identifiable parts.

If the ether had identifiable parts then differences in *C* (if such were found) might be capable of unambiguous description in terms of motion relative to the ether; but the ether has baffled all attempts to identify it.

The existence of a unique ether would imply the variation of *C* with the motion of the observer—whether this motion were uniform or accelerated—but as experiment has shown that no such variation occurs we are driven to conclude tentatively that a unique, identifiable ether does not exist.

The velocity of light relative to the observer is, however, a quantity which enters into the description of electromagnetic phenomena and so, at least indirectly, into that of all phenomena, and it would be most difficult to conceive of any law of variation of this velocity which would make an unambiguous description of phenomena possible without confusing complexity, except the law that the variation is nil. For instance, unless such were, at least approximately, the law, the whole structure of description known as the science of electromagnetism would be wrong.

We expect, therefore, to find that *C* in free space is constant because we find that science has long been fairly orderly and fairly free from ambiguity.

The TO
diagram

We return to the Special Theory.

In this theory the velocity of light is constant in relation to all possible observers, for the theory deals only with uniform motion in regions free from gravitation.

We give to each of two observers *S* and *S′* a standard metre rod. This serves both as a rod and as a clock—in the manner already described—the passage of light back and forth from one end of the rod to the other a definite number of times representing one second.

S observes a particle and finds it to be at rest relative to him. He might plot the space-time "path" of this particle as in Fig. 52. Its "path" (world-line) is *OT*. *S* observes also a ray of light starting out from the point in space which *S* regards as a spatially fixed starting point, at the instant in time which he regards as zero time.

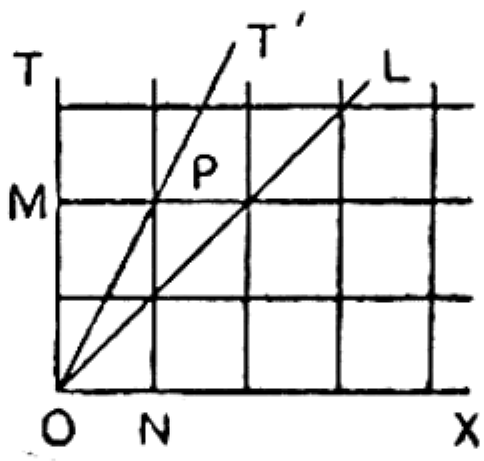

FIG. 52

If the line *OT* is divided into equal parts representing units of time (seconds) and *OX* into equal parts representing each one unit of length (186,000 miles) then the "path" of the light ray will be the line *OL* inclined at 45 degrees to *OX*. Following the usual convention *S* makes his space axis *OX* perpendicular to *OT*, since, otherwise, the "path" of a particle "at rest" (that is to say "progressing along" *OT*) would have a projected component along *OX*, as though it were not at rest but in motion.

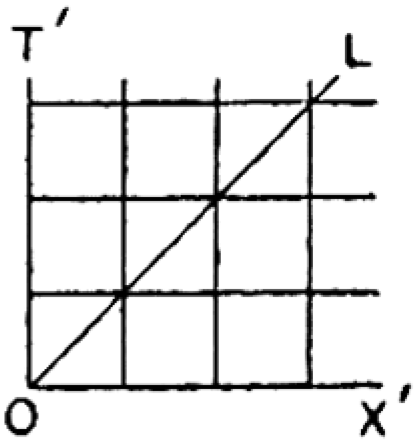

FIG. 53

The other observer *S*′, in uniform motion relatively to *S*, observes the same ray of light and, reasoning in the same manner that *S* did, *S*′ makes a diagram (Fig. 53) similar exactly to that made by *S* (Fig. 52), the axes of which must be labelled, however, *OT*′ and *OX*′ to show that they refer to *S*′ and not to *S* space-time.

But to superimpose the two diagrams brings confusion. *OT* represents *S* path relative to himself and *OT*′ represents *S*′ path relative to *himself*, and to superimpose *OT* and *OT*′ is inconsistent with the fact that *S*′ is in motion relatively to *S*. Similarly inconsistent is the superposition of *OX*′ upon *OX*.

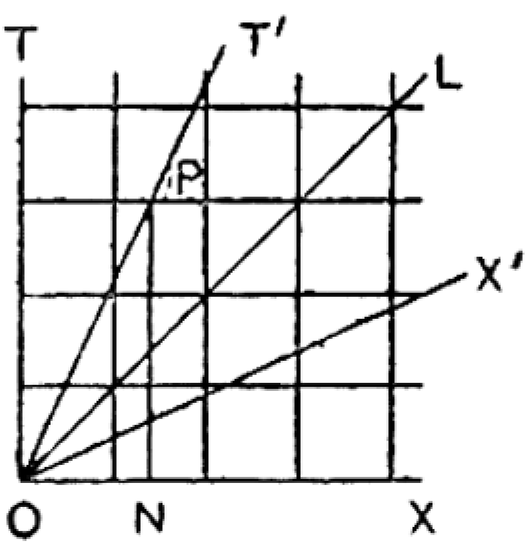

FIG. 54

A space-time diagram which shall truly represent both the relative motion of *S* and *S*′ and also the motion of the light ray relative to both observers must be one in which *OT*, *OT′*, *OL* and *OX* and *OX′* are as in Fig. 54, the slope of *OT′* having been chosen such that *ON*/*NP* (measured in space units and time units respectively) equals the velocity of *S′* relative to *S*.

The instructions for drawing *S′* space axis *OX′* are contained in the fact that the velocity of the light ray whose path is *OL* must be unity in *S′* frame as well as in *S* frame; that is to say, provided we may assume that *S′* length and time units are equal just as we have made *S* units, the line *OL*, being equally inclined to *S* axes, must be equally inclined to each of *S′* axes also.

In regard to the equality of *S′* time and space units this is ensured, just as it was for *S*, by equipping *S′* with a proper rod (to act also as his clock). *S′* is justified in assuming that no change has occurred in his rod because he is justified in maintaining that *he himself is at rest*. *S* says *S′* is moving with velocity "*u*" and that he himself (*S*) is "at rest," but *S′* is equally well warranted in saying that *S* is moving with velocity "*u*" while he *himself* (*S′*) is "at rest."

We see then that in Fig. 54 *TOX* is *S* frame, and *T′OX′* is *S′* frame, the former using rectangular meshes and the latter using oblique ones. There is no "real" difference between these "differently shaped" meshes because shape is not real. If, for instance, we had happened to deal first of all with *S′* instead of with *S* we should probably have given the rectangular meshes to *S′* and, calling him stationary, should have attributed the velocity "*u*" to *S*. Both *S* and *S′* may be thought of as starting together at *O* which, if we like, may be pictured as a wooden chair on a long straight concrete path *OX*. *S* remains in the chair while, at a given instant, *S′* and a light ray start off along the path *OX* the former with velocity "*u*" (relative to *S*) and the latter with velocity unity (= 186,000 miles per sec.) relative to *S* (and equally, we may say, relative to *S′*).

Thus the space-path of *S* is the point *O*, that of *S′* and of the ray is *OX*, but the space-time paths are, *OT* for the stationary man S, OT' for S' (the obliquity of OT' being such that $ON/PN = u$), and OL for the ray.

We have now to discuss the important question of *the relative values of the units that must be used by S and S'*. The discussion will lead in a very simple way to the following conclusions:

1st. That S' units are greater than those of S in the ratio

$$\sqrt{(u^2 + 1)/(1 - u^2)}.$$

2nd. That S' rod contracts and becomes equal in length to $\sqrt{1 - u^2}$ times S rod.

3rd. That S' clock goes more slowly than S clock, the number of ticks it makes in a given time being $\sqrt{1 - u^2}$ times the number of ticks made by S clock.

4th. That S' clock is behind that of S in its setting, the difference $(x'u/\sqrt{1 - u^2})$ of S' units, depending on the position of S' clock in S *space*.

Afterwards we shall express these results in the form of a set of equations called the Lorentz Transformation equations, and when from this transformation we have deduced the effects of gravitational fields upon rods and clocks, effects which, as we have seen already, contain in themselves Einstein's law of gravitation, we shall have succeeded in deducing *without reference to experiment* the chief propositions of Einstein's General Theory of Relativity.

Throughout the discussion attention should be kept upon the fact that *assumptions* well warranted by a mass of old experimental facts, and not new experimental facts, are the basis of the argument. These assumptions are, mainly, the following:

(1) That the velocity of light as measured locally with rod and clock is found to be always the same, no matter how the observer may be moving relatively to the source of light, or in what position in "space" (gravitational or not) he may be.

(2)—(Not strictly a new assumption because its truth is demanded by the latter part of (1) relating to gravitational fields) —The *Principle of Equivalence* which postulates that in any small region of space a gravitational field is experimentally indistinguishable from an acceleration of the observer's frame of reference.

(3) The principle of the *relativity of motion* which postulates that it is impossible to measure the motion of a particle without reference to some other particle.

In regard to (1) we have indicated already how the constancy of light's velocity is almost demanded by the general state of science. Assumption (2) becomes almost a "self-evident truth" when we consider the mass of evidence there is in favour of the statement that freely-falling bodies experience no gravitation. To allow a body to fall freely is simply (and, so far as we know, is solely) to give it an acceleration, and since, invariably, this acceleration neutralizes gravitation, we conclude that it is equivalent to gravitation.

Lastly (3) the principle of the relativity of motion has to be accepted because—taking in turn all the factors there are in "motion," namely, acceleration, velocity and position—first, acceleration cannot be measured except in terms of velocity, secondly, velocity cannot be measured except in terms of position, and, thirdly, position cannot be measured except relatively.

In discussing the question of the relative values of S and S' units we shall refer only to the TOX diagram. In this diagram —referring to Fig. 54—we have seen that the condition that both S and S' shall find light's velocity equal to unity has been satisfied. The next point is to ensure that S and S' find each the same numerical value "u" for the other's velocity.

Velocity is defined as distance divided by time and if, for instance, in the oblique mesh-system of Fig. 55 a particle should move, say from O to P, its velocity must be expressed as (space units/time units) $= ON/PN = 2/4$.

We refer next to Fig. 56 in which the time and space axes

OT and *OX* are shown for *S*, and *OT′* and *OX′* for *S′* (moving with velocity $u = ON/PN$ away from *S*).

We have to show that the velocity of *S* relative to *S′*, as judged by *S′* and measured in *S′* own units, has the same value *u* which that of *S′* has as measured by *S*. This means that we have to show that the slope (ON/PN) of *S′* path *OT′* in *S* frame *TOX* is the same as the slope (RP/OP where *RP* is parallel to *OX′*) of *S* path *OT* in *S′* frame *T′OX′*.

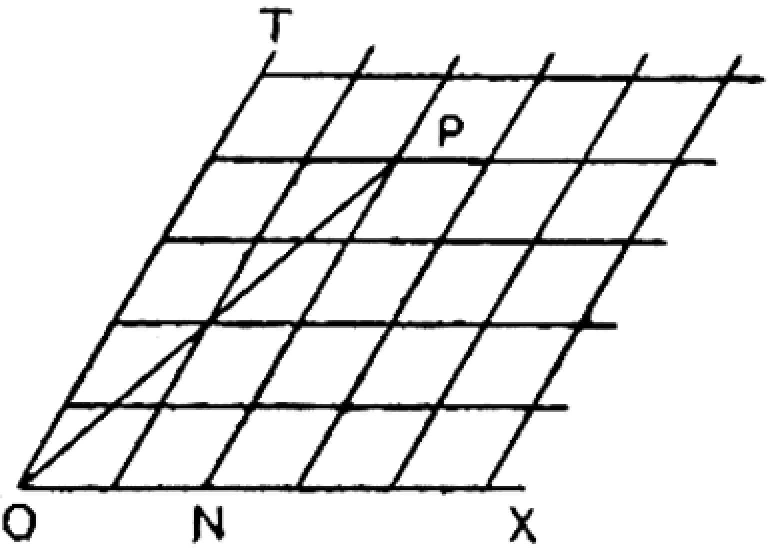

FIG. 55

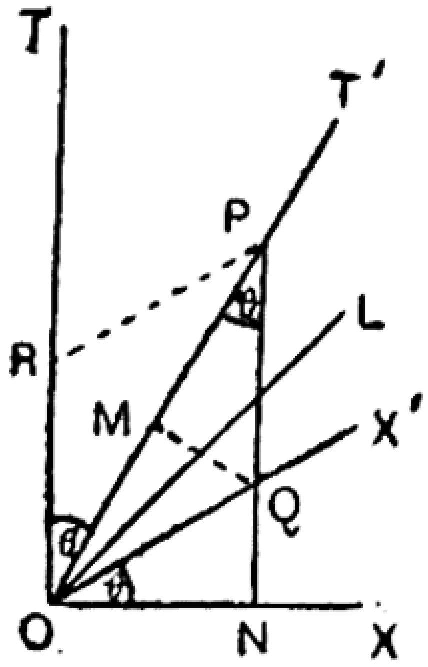

FIG. 56

Drawing *QM* and calling the angle *TOT′* θ, we have

$$\cos\theta = ON/OQ = PN/OP, \quad \therefore \; OQ/OP = ON/PN.$$

But $OQ = RP$, hence the slope RP/OP of *OT* in $T'OX' =$ the slope ON/PN of *OT′* in *TOX*.

It will be noted that this result holds, no matter what may be the magnitude of the units used by *S′* relatively to the magnitude of those used by *S*.

This question of the relative magnitudes of the units is answered when we express the fact that the percentage change which each observer remarks in the other's rod and clock shall be numerically the same.

In Fig. 57 only *S′* meshes are shown, those of *S* can be imagined.

We have to find out (referring to Fig. 58) what must be the ratio of the paper distance l' inches representing one of *S′* length units to the paper distance l of an *S* length unit, in order to fulfil the condition that the numerical value of the percentage change which either S or S' sees in the other's clock shall be the same as that of the change he sees in the other's rod.

It will simplify the discussion if we first draw Fig. 58 in which the units chosen actually do fulfil approximately the required conditions, and explain, by reference to this figure, just what is meant by S and S' "opinion" of the behaviour of the other's rods and clocks.

The argument is simple:

S is observing and he sees S' rod OP (at rest in $T'OX'$) passing through S' *time*, and in so doing traversing S frame (along the channel bounded by OT' and RP) and S sees that S' rod, thus obliquely projected to disclose its "S-space" length, is OR and is less than one unit of S-space OQ. Thus S says S' rod has contracted.

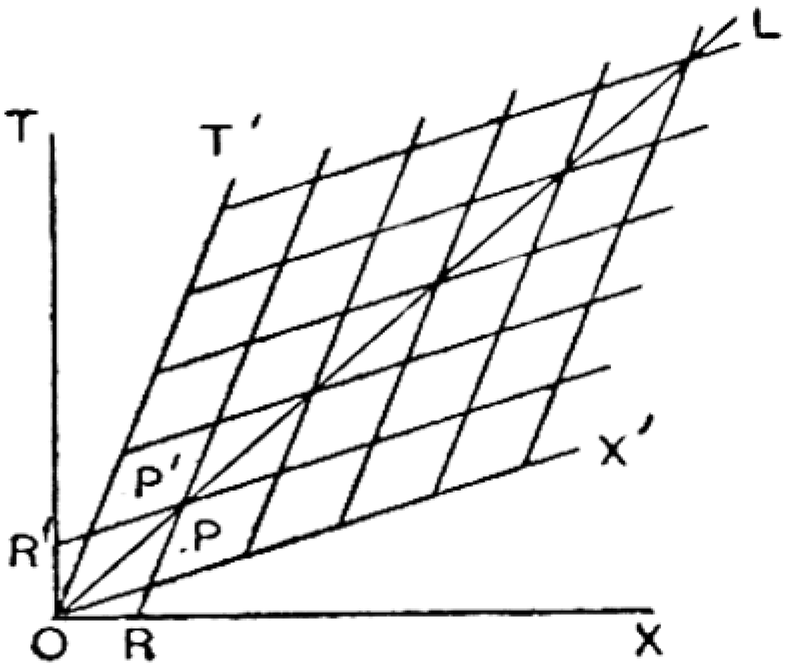

FIG. 57

The ticks of S clock occur in S' frame and in this frame two S ticks (O and Q' in Fig. 58) are more than one S' time channel width OR' apart. Thus S' says S clock goes slowly.

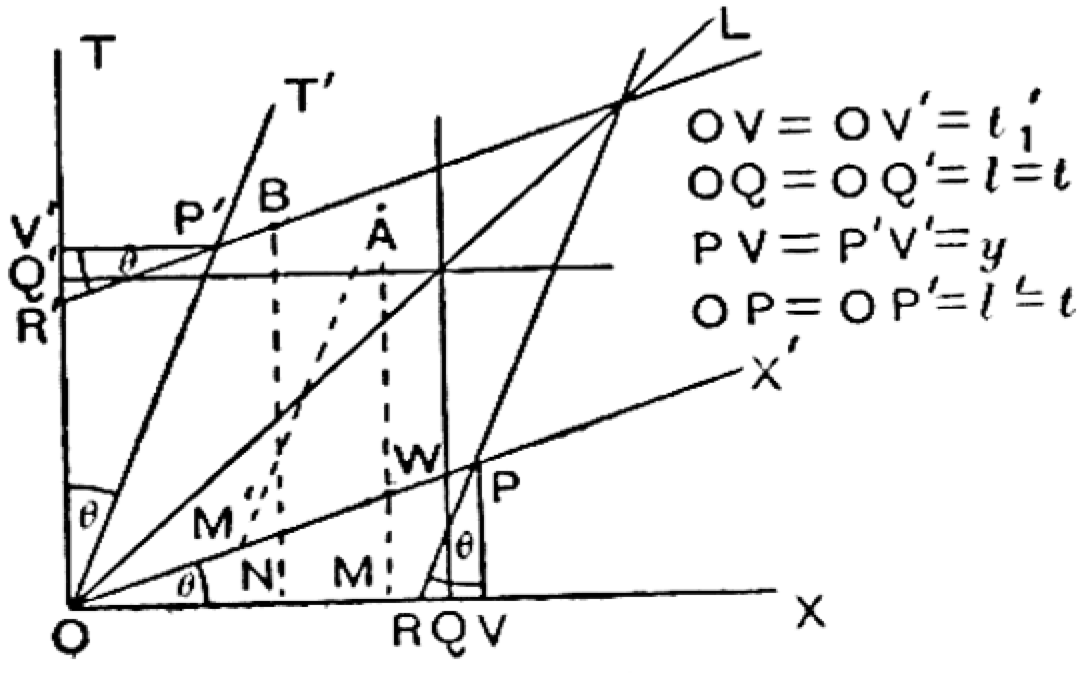

FIG. 58

In order to become more accustomed to the TOX diagram we next consider S' point of view in regard to S rod and clock. We shall find that the points of view of S and of S' are reciprocal. S' is observing and he sees S rod OQ (at rest in

TOX) passing through *S* time and in so doing traversing *S′* frame (along the channel bounded by *OT* and *QW*) and *S′* sees that *S* rod, thus obliquely projected to disclose its "*S′*-space" length, is *OW* and is less than one unit of *S′*-space *OP*. Thus *S′* says *S* rod has contracted.

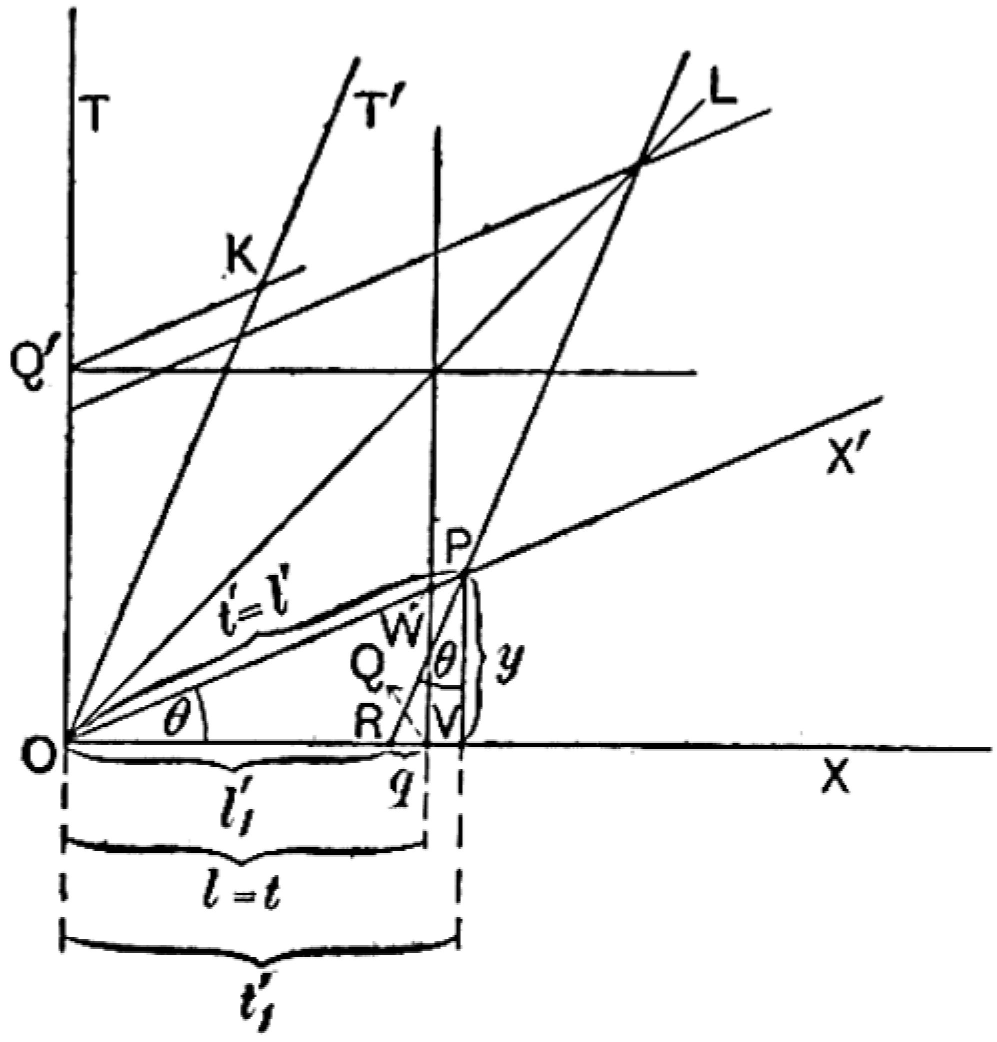

FIG. 58 A

Again: The ticks of *S′* clock occur in *S* frame, and in this frame two *S′* ticks (*O* and *P′*) are more than one *S*-time channel width *OQ′* apart. Thus *S* says *S′* clock goes slowly.

Geometrically the problem is to express the condition that *S* opinion of *S′* time unit is to *S* time unit as *S* length unit is to his opinion of *S′* length unit.

Referring to Figs. 58 and 58 A

S	opinion of	*S′* time unit	t' $(=OP')$	is that its	value is	t_1' $(=OV'=OV)$
S	,,	*S′* length ,,	l' $(=OP)$	,,	,,	l_1' $(=OR)$
S′	,,	*S* time ,,	t $(=OQ')$	,,	,,	t_1 $(=OK^*)$
S′	,,	*S* length ,,	l $(=OQ)$	,,	,,	l_1 $(=OW)$

Thus we have to find what must be the value of the ratio t'/t $(= l'/l)$ in order that

$$\frac{t_1'}{t} \text{ shall be equal to } \frac{l}{l_1'}.$$

We attack the problem by assuming that this desired condition holds in Fig. 58, that is to say we assume that

$$\frac{t_1'}{t} = \frac{l}{l_1'} \text{ or (since } l = t\text{), } \frac{t_1'}{t} = \frac{t}{l_1'} \text{ or } t = \sqrt{l_1' t_1'}.$$

---

* Where *K* is a point on *OT′* such that *Q′K* is parallel to *OX′*.

Then in the figure, $(y/t_1') = \tan\theta = u$, the relative velocity of S and S'.

$\therefore\ y = ut_1'$.

And in the triangle OVP

$$y^2 + t_1'^2 = t'^2. \quad \therefore \quad (ut_1')^2 + t_1'^2 = t'^2,$$

and

$$t' = \sqrt{(ut_1')^2 + t_1'^2}.$$

Then

$$\frac{t'}{t} = \frac{\sqrt{(ut_1')^2 + t_1'^2}}{t = \sqrt{l_1't_1'}} = \sqrt{\frac{u^2t_1' + t_1'}{l_1'}}.$$

Now express l_1' in terms of t_1' thus:

$$l_1' = t_1' - q \quad \text{and} \quad (q/y) = \tan\theta = u. \quad \therefore \quad q = yu = u^2t_1',$$

thus $l_1' = t_1' - u^2t_1' = t_1'(1 - u^2)$, so that, finally,

$$\frac{t'}{t} = \sqrt{\frac{u^2 + 1}{1 - u^2}} = \text{also } \frac{l'}{l}.$$

Thus the number of inches between consecutive divisions on S' axes of length and of time is equal to $\sqrt{(u^2 + 1)/(1 - u^2)}$ times the number of inches between consecutive divisions on S axes of length and of time. This means that S' units *both of length and of time* are of greater paper-length than S units.

It is now easy to show that, on this account, S and S' are each of the opinion that the other's rod has *contracted* so as to become only $\sqrt{1 - u^2}$ of a *proper* unit long, and that the other's clock is going slowly so that its seconds are $(1/\sqrt{1 - u^2})$ times a proper unit of time.

We first find t_1'/t thus:

$$t_1'/t = (t_1'/t') \times (t'/t) = \{t_1'/(t_1'\sqrt{u^2 + 1})\}\,(t'/t)$$

because

$$y^2\,(= u^2t_1'^2) + t_1'^2 = t'^2,$$

hence

$$(t_1'/t) = (1/\sqrt{u^2 + 1})\,(t'/t) = (1/\sqrt{u^2 + 1})\sqrt{\frac{u^2 + 1}{1 - u^2}} = \frac{1}{\sqrt{1 - u^2}},$$

a quantity greater than unity often called β.

Thus finally

$$(t_1'/t) = (l/l_1') = (1/\sqrt{1 - u^2}) = \beta.$$

The relative setting of clocks

The *TOX* diagram alone does not disclose the amount of the relative *setting* of *S* and *S′* clocks. This is because in this diagram the whole of the difficulty met with in practice (and resulting, as we shall see, from the peculiar property of light's velocity that it is the maximum possible velocity with which any body or influence can travel) of defining simultaneity in time at different points in space is not disclosed. Thus, for instance, in Fig. 58 (p. 215) the point *P′* in space-time is called 1 o'clock, say, by *S′* and is called (*OV′*=), say, 1^{15} by *S*, that is to say the points *P′* (1 o'clock to *S′*) and *V′* (1^{15} to *S*) are *simultaneous*. But this means simultaneous from *S* point of view, not from that of *S′*, for to *S′* the point *R′* is the one which is *simultaneous* with *P′*. Thus the *TOX* diagram does bring out the relative nature of time estimates, but only in so far as this depends on the differences of units used by different observers.

The *TOX* diagram postulates that all points along *OX* are simultaneous to *S*, and all points along *OX′* are simultaneous to *S′*.

But we said *OX* might be pictured as a straight rigid concrete road with a wooden chair fixed at *O*, for *S* to sit in. *S* was to call distance along that road *pure space*.

To be fair we now give *S′* a seat also, but it must be a seat in, say, a motor-car moving with velocity *u* along the road *OX*. Then *S′* regards the road *OX* not as pure space but as partly time. *S′* knows the road is not pure space because it occupies both space and time—he travels along it and finds by experience that *OX* involves time as well as space. To *S′* pure space is *OX′*, not *OX* at all. But why *OX′*? Because this is the only space path (in the particular 2-dimensional space-time continuum represented by *TOX*) in which *S′* seat (and motor-car) are at rest.

Similarly *S* calls *OX* pure space because in *OX* he and his seat are *at rest*.

In fact our rule, when we speak as individual observers, is to give the name "pure space" to *any path along which we are* *not travelling*. But is this logical? What do we *know* about a path unless we examine the whole of it? Does it even exist? We reply that we have other means of examining a path besides following it in person. We may pass a hand along it, or trace it with a long stick, or in extreme cases we may just *look along it*. But does not this mean, in each case, to send some sort of a moving thing or influence along the path; since even a beam of light has to travel?

A thing which *travels along* the path must report, however, that the path is not *pure space*, because experiment showed that it occupied both space and time.

It is hard for common-sense to have patience with such a line of reasoning. Our reply is obvious. It is true that we can *name* no thing or influence which could travel along that road *in no time*, but we can *conceive* of one, and that satisfies us. When we draw *TOX* diagrams (which are essentially devices for illustrating the Newtonian conception of space and time) the line *OX* represents the space-time path of a traveller having *infinitely great velocity*. It is just at this point that the Relativist parts company with the man of common-sense: he refuses to be satisfied with a *conception* unless it is forced upon him by experience. The truth is that which *works*—not that which we conceive ought to work.

Let experimental evidence be found of the existence of some thing or influence which has infinite velocity: the Relativist will concede at once that his postulate of the relativity of space and time may possibly have to be withdrawn, the General Theory of Relativity thus taking its place as a stepping stone to some still more inclusive theory, much in the way that the Special Theory was a stepping stone to the General.

So far, however, as experience has yet gone, we may state confidently that *no velocity greater than that of light exists* in nature. Relativity proposes to assume, tentatively, that no greater velocity ever will be found to exist.

It *is* a fact that any given *TOX* diagram represents, *not* reality, but just one individual point of view.

Bearing all this in mind we now follow the observer S' in his car. He moves away from S along the concrete road OX—a path which to him (S') is partly time. Both S and S' are recording their observations of, say, a moving aeroplane, and they have agreed to make *truly simultaneous* observations. They have similar clocks and rods (calibrated together at O while both S and S' were together there and at rest) and although they realize the difficulty about the effect of motion upon rods and clocks, and have worked out the formulae by which to correct their readings each to the other's point of view, this is not enough. It is true they might arbitrarily *define* their own idea of simultaneity—by saying, for instance, that always the instant at which one observer's clock said 12 o'clock should be simultaneous with the instant at which the other's read 12 o'clock (or whatever time t' when multiplied by β gave a time equivalent to 12 o'clock). But examination of the TOX diagram shows that, owing to the slope of $T'OX'$ relative to TOX, this definition would be consistent with neither S nor S' conception of space and time, nor would it agree with the ideas of other *independent* observers.

Some means of setting and re-setting clocks from time to time is demanded if the records of a large number of independent observers are to be taken as a basis for a consistent scheme of description of events.

S and S' decide, therefore, that they will signal to each other from time to time by means of flashes of light, but, realizing that "from time to time" is perhaps ambiguous, they say rather that they will signal at various "places." For instance, at the (S) instant when S signals, say, the instant "noon" to S', S will measure his own distance from S' and will signal to S' the value of this distance (expressing it, for S' convenience, in S' units). They agreed beforehand that S should send this "place signal" exactly when his (S) clock read "noon." There is no ambiguity about their relative distance apart, so far as merely S and S' are concerned, for each knows that the other's rod contracts to $(\sqrt{1 - u^2})$ times its "proper" value. For instance, if S finds S' to be distant x units away (which for clearness we will designate x_s units, the suffix S signifying that the numeral x means S units) he knows that this may be written (using the symbol β to stand for $(1/\sqrt{1 - u^2})$, a quantity *greater than* unity) $(x'/\beta)_s$—still in S units. Then as they are in agreement as to the value u of their relative velocity, and also as to the value unity of C, each observer can correct for the time taken for the signal to reach him.

S knows that if S' is to make his clock truly synchronous with S clock then, on looking (say by telescope) at S clock, S' ought to *see* it reading rather *less* than his own (S') clock reads, because the light has taken time to come from S to S', the time required being $\{x/(1 - u)\}$ of S time units. S knows that S' is moving with velocity u, and also he knows, by direct experiment, that light signals leave himself (S) with velocity C ($=$ unity), hence he concludes that the velocity with which a light signal overtakes S' must be $(1 - u)$: then as x_s is S' distance away at the instant when S sends the signal, S knows that, in order to overtake S', the signal requires $\{x/(1 - u)\}_s$ units of time, or $\{(x'/\beta)/(1 - u)\}_s$ still in S units of time. This time interval may be written

$$\{(x'/\beta^2)/(1 - u)\}_{s'}$$

in S' units of time, and may be put into the form $(x' - ux')_{s'}$ in S' units of time.

Thus S says that if, on receipt of the "noon" signal (telling S' that he is $x'_{s'}$ units of distance away), S' will set his own (S') clock at {noon plus $(x' + ux')_{s'}$}, then S' clock will be "correct."

But S' has measured the velocity with which light signals approach himself and has found it to be unity; hence he *knows* that the signal which S has sent when distant $x'_{s'}$ away must have been x' units of time; he therefore *knows* that his own (S') clock will indicate "correct" time if he sets it to read {"noon" plus $x'_{s'}$} at the instant when the "noon" signal arrives.

Thus S expects S' to put his clock $(x' + ux')_{s'}$ units ahead of noon and actually S' puts it only (x') units ahead; therefore S' clock is *backward* by the amount $(ux')_{s'}$ in respect to S clock.

Hence in order to make S' "time" equal to $(1/\beta)$ of his own time (as he knows it must be since experiment tells him that S' clock goes slowly) he has to regard S' "time" as the reading $t'_{s'}$ of S' clock, plus $(ux')_{s'}$; he thus writes

$$t/\beta = (t' + x'u)$$

or $t = \beta\,(t' + x'u)$ where t stands for the reading of S clock as read by S and t' for that of S' clock as read by S', while x' is the distance between S and S' in S' units. The equation tells us that S and S' have different ideas about the meaning of time and of space. From S' point of view (and it is just as legitimate as the point of view of S or any other observer) "time" is the thing indicated by his own clock, and t' is its "true" value, but from S point of view a factor $(x'u)$ has to be added to t' to turn it into "true" time.

Taking now the argument from S' point of view we need only interchange the dashed and the undashed symbols throughout and replace $+u$ by $-u$. S and S' both give the same numerical value to their relative velocity, but since S calls it positive S' must call it negative. Thus the equation for t' is

$$t' = \beta\,(t - xu)$$

which shows that from S' point of view a factor (xu) has to be subtracted from t to turn it into "true" time. It should be noted that if the symbol C be used for the velocity of light (which in our chosen units is unity) the equations contain (xu/C^2) instead of xu, the fact thus becoming explicit that the xu term is of the nature of a velocity.

An alternative method of deducing the setting adjustment of moving clocks

S, "at rest," carries with him a source of light (O), a beam from which is overtaking S' who moves away from S with uniform velocity u in the direction Ox.

If *rods and clocks were unaffected by motion* and if *light obeyed Newton's rule of relative velocity* we could say confidently that, on measuring the velocity with which the light beam overtook him, S' would obtain the value $(1 - u)$, and would find that the beam covered a distance x' in time $t' = x'/(1 - u)$.

We know, however, that S' rod contracts and that on this account S' must be expected to find light's velocity $\beta(1 - u)$ and the time $x'/\beta(1 - u)$, while, in addition, his clock goes slowly so that S' must be expected to find a still further increase in the velocity, bringing it finally to $\beta^2(1 - u)$ and the time to

$$x'/\beta^2(1 - u) = x'(1 - u)(1 + u)/(1 - u)$$
$$= x'(1 + u) = (x' + u'x).$$

But according to Relativity S' must find a velocity equal to unity and a time equal to x'. We have to suppose, therefore, that the clocks which S' puts, one at each end of the path x', must be set the one backward in respect to the other, the more distant one being backward by ux'. Thus S' uses a time interval t' which, from S point of view, is too small and S adds ux' to it before he writes the equation

$$t = \beta(t' + ux').$$

Time order may be different to different observers

Inspection of the *TOX* diagrams discloses a curious possibility, upon which the relative rates or settings of clocks have no bearing, namely that the time *order* of the occurrence of one and the same series of events may be different to different observers. This means that there is no "reality" about time order and makes necessary a readjustment of our point of view in regard to "cause and effect," since the idea that under any possible circumstances an effect could precede its cause is inconsistent with the conventional point of view. The consideration emphasizes the fact that we have to take things as we find them, all we can hope to do being to describe them.

Referring to Fig. 58 in which *TOX* is the frame of one observer S whom we may regard as "at rest" and $T'OX'$

that of *S′* moving with a velocity of about ½ (that is to say about 93,000 miles per second) relatively to *S*, we consider the "events" *A* and *B*. In *S* frame the time co-ordinate *AM* of *A* is longer than the co-ordinate *BN* of *B*; hence, to *S*, *B* happens before *A*. In *S′* frame *T′OX′*, however, the time co-ordinate (= *P′O*) of *B* is longer than the co-ordinate *AM′* of *B*, wherefore, to *S′*, *A* happens before *B*.

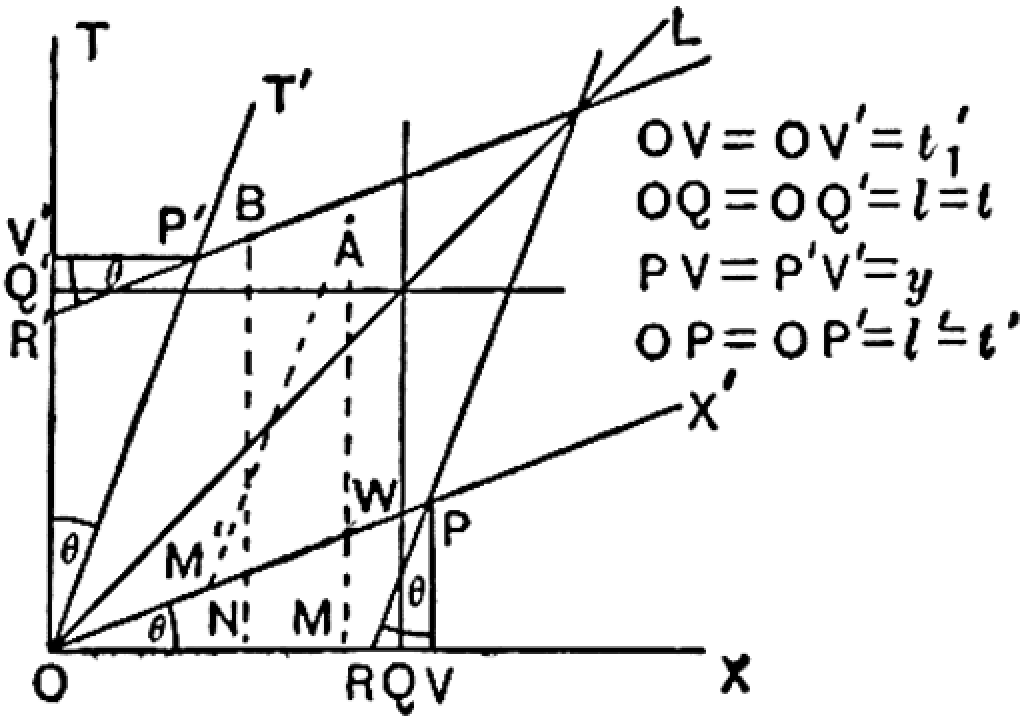

FIG. 58

The FitzGerald contraction in the light of common-sense

The FitzGerald contraction and its consequence, the slowing of clocks, are found to be a stumbling-block to those who wish to follow the Theory of Relativity yet cannot give up their allegiance to common-sense. They have the feeling that such a contraction must cause serious troubles—such, for instance, as railway accidents—if it *really* occurred. Careful consideration of special cases will show, however, that whereas *transverse* changes in dimensions would cause practical catastrophes, changes, as postulated by Relativity, in the direction of motion cannot. The FitzGerald contraction represents a change in dimensions which not only is real, but which cannot create or destroy contacts between material particles—it is a percentage reduction in distances, never an annihilation of them.

Transverse changes of length cannot occur

Many practical examples may be chosen—as, for instance, examples of trains travelling side by side with given relative velocity and having arms upon them, which can be made to protrude and be suddenly withdrawn, and so on—but if ever reasoning should show that the FitzGerald contraction must cause the creation or destruction of contacts, then that reasoning is surely wrong.

Contacts are space-time coincidences, and these are "real" events, whereas the FitzGerald contraction is a mere change in length, that is to say a change in something which characterizes the relation of some observer to some events, and which does not belong to the events at all. One specifiable observer will see the contraction; to another it will not occur; hence if it created a collision this collision would be a thing which appeared to the one observer and not to the other and so the collision would be unreal.

A tempting sort of example to choose in the hope of showing that the FitzGerald contraction would cause accidents if it "really" occurred is one involving electrical contacts. We arrange, for instance, a pair of contacts 100 ft. apart on the rails of a railway and provide corresponding contact arms on the train, also 100 ft. apart, in such a way that when they strike the rail contacts *simultaneously*, some device will be actuated, the train brakes being caused, for instance, to be applied to the wheels.

The suggestion then is that if the FitzGerald contraction did *actually* occur to the moving train then the arms upon this train would become less than 100 ft. apart, the electrical circuit would not be completed, the brakes would not be applied, and an accident would occur.

In Fig. 59 *P′* and *Q′* are the contacts upon the train, *P* and *Q* those upon the rails and *S* is the observer. *S* finds that *P′Q′* are less than 100 ft. apart owing to the FitzGerald contraction caused in the train by its motion. Hence at the instant when *P′* reaches *P*, *Q′* will not yet have reached *Q*, and the electrical circuit via *PQQ′P′* will fail to be made because *Q′* strikes *Q* *too late*.

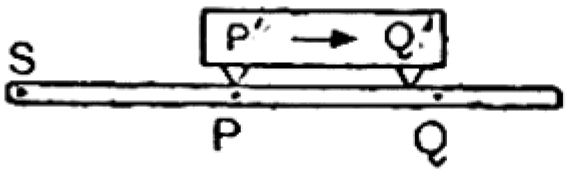

FIG. 59

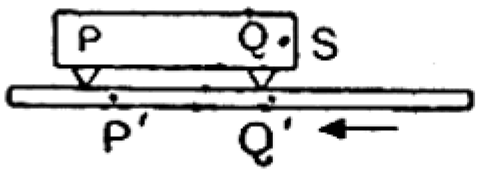

FIG. 60

Here we have regarded the rails and the observer at *S* upon them, as at rest, but it is equally legitimate to regard the train as at rest and the rails as in motion.

In Fig. 60 the point of view is taken of the observer *S* at rest on the stationary train watching the rails with their contacts *P′* and *Q′* as they move beneath the train. The observer *S* sees that *P′Q′* is shorter than *PQ*, so that when *Q′* strikes *Q* the contact *P′* has still some distance to travel to reach *P*. Thus from the train observer's point of view the circuit fails to be made because *Q′* strikes *Q* *too soon*.

One observer says the space-time coincidences *PP′* and *QQ′* occur in the time order *PP′* – *QQ′* while the other says these same "real events" occur in the time order *QQ′* – *PP′*. Is it surprising that neither description is right? That which in "reality" happens is that the contacts occur in the one definite space-time order which ensures the passage of the current.

Common-sense is helped by looking at the matter in the following way:

Instead of dwelling upon the delay of *Q′* in reaching *Q* (Fig. 59) due to the contraction of *P′Q′*, we think only of the contact arm *Q′*. This, being fixed to the train, travels with velocity *u* relative to *S*, hence the time which elapses, in *S* estimate, between the crossing of *P* by *Q′* and the arrival of *Q′* at *Q* is (x/u), since x is the distance *PQ*.

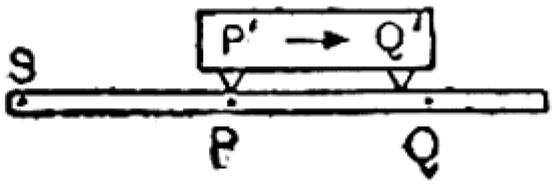

FIG. 59

Now with rails and train relatively at rest the distance *P′Q′* had been made originally equal to *PQ*, and this ensured that if *Q′* reached *Q* in (x/u) second after *Q′* crossed *P* the contacts would be made *so as to allow current to flow*.

The fact that *to S* the contacts *Q′Q* and *P′P* are not simultaneous does not matter, simultaneity being a relative thing. The contacts are the realities, and we have to ensure that these occur in such a way as to make a continuous circuit through *P′PQQ′*. This has been ensured independently of any observers, by making *P′Q′* coincide with *PQ* when both train and rail are at rest.

We may take still another point of view. Let the contact points *P′Q′* be *at rest* upon the points *PQ*. The circuit is thus complete. What happens if the train now suddenly obtains (without appreciable change of position) a velocity *u*? The train is now sliding over *PQ*, but as (at the very instant of contact) such motion is transverse and therefore, as will next be shown, not *relative* motion of *P′Q′* in respect of *PQ* there is no reason why *P′Q′* should contract relatively to *PQ*. The attainment of the velocity *u* does not, therefore, break the circuit *P′PQQ′*.

Transverse motion

This raises the question of the meaning of transverse motion. According to Relativity, motion is meaningless except in relation to some definable thing (frame). It is only when a particle is changing its distance from some other particle that there is any sense in saying the particle is in motion.

What is the meaning of saying a particle *P* is in transverse motion—purely transverse motion—relative to another particle *O*? The meaning can only be that, at the instant considered, the direction of *P*'s motion is perpendicular to *OP*. And if the pure transverse motion is to continue then *P* must move in a circle about *O* as centre.

Thus the question of the meaning of pure transverse motion is that of the meaning of rotation. When *P* moves in a circle about *O* its distance from *O* does not change; hence, according to Relativity, there is no sense in saying that *P* is in motion relatively to *O*. Hence pure transverse motion is without meaning.

Thus when, with *P′* and *Q′* at rest and in contact with *P* and *Q*, a sudden velocity is given to the train there is no sense in saying that *P′* is thus given a velocity relative to *P* or *Q′* relative to *Q*, and no sense in supposing that the mere *possession* of velocity by the train can break those contacts.

These considerations suggest still another way of looking at Fig. 59—a way which, it is hoped, may make the nature of the difficulty quite clear.

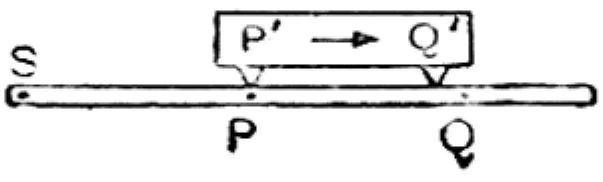

FIG. 59

The difficulty introduced by the FitzGerald contraction occurs in respect only to the train as a whole—the contraction having the effect of drawing Q' back towards P'—not in respect to the motion of P' or of Q' or of any spot on the train when this spot is regarded merely as a moving particle. P' and Q' and every other particle on the train move relatively to S with one and the same unambiguous velocity u; moreover, the particles are in their proper succession; hence, since the electric current is conducted from particle to particle there is no reason to suppose that the progress of the current can differ in any way from that which S would have expected if he had never looked at the train or been worried by the possibility of the FitzGerald contraction. The reason why, when looking at the moving train, S sees a contraction which makes Q' reach Q too late, is that S insists in dragging into his observation the irrelevant matter of the readings of clocks —clocks which he likes to imagine attached to P' and to Q' and to every other moving particle. For instance: S has calculated that contact at PP' will occur at 12 noon precisely and if "simultaneous" contact is to occur at PP' and at QQ', the QQ' contact must occur also at 12 noon; that is to say S expects to see Q' reach Q precisely when it is "really" 12 noon. What he does see, however, is that Q' reaches Q precisely when the clock carried on the particle Q' *reads* 12 noon—and as S considers that the clock at Q' reads *backward* S concludes that the real time is past 12 noon, that is to say he concludes that Q' has arrived at Q too late.

It may be asked why did not S put a clock at Q (at rest relatively to himself and so quite synchronous with his clock at S) to compare directly with the clock at Q'? If he had done so we know for two simple reasons that the Q' clock and the Q clock both must have indicated 12 noon when Q and Q' came into contact. The one reason is that, as already explained, when S puts attention simply on the particle Q', travelling with velocity u, he knows that it must reach Q at the instant at which he expects it to do so (say, 12 noon), while with equal legitimacy and "truth" an observer on the train knows that Q will reach Q' exactly when the Q' clock reads 12 noon. The other reason is that if, having placed a clock at *Q*, *S* now takes up his own position at *Q*, then when *Q'* is sliding over *Q*, *S* will observe that *Q'* has no velocity relative to *Q* and, therefore, that his own and *Q'* clock both are in agreement and are going at equal rates.

The reply to these very interesting criticisms is this, that if every observer were indeed in the fortunate position of *S* of being able to put one of his own clocks, at rest relatively to himself, at every point in the space path of every moving particle there would be no need for the Theory of Relativity. When we postulate that *S* is able to put at any and every point on the rails, clocks "truly" synchronous with his own, we are postulating that *S* has knowledge such as in fact no observer can possess, namely, knowledge of the meaning of simultaneity in time at different points in space.

It is our attribution to the rails of the mysterious and indefinable property of rigidity which leads us to assume what in fact cannot be proved, namely, that by means of rigid connections from point to point in space, simultaneity in time can be defined unambiguously. If that which one observer calls rigid appeared so to all other conceivable observers the assumption would be warranted, but the Michelson-Morley and other experiments have demanded the FitzGerald contraction and this makes rigidity a relative matter.

The consideration of transverse changes in the dimensions of moving bodies is simple. Had the Relativist postulated a *transverse* change in dimensions, due to motion, we could easily have shown him to be wrong by citing instances in which such an effect would produce accidents or new contacts. One such instance may be given. A train moving along rails and obeying all the ordinary rules of Newtonian mechanics, suddenly becomes the object of attention of an observer fixed to the rails some distance away. If it be postulated that the motion of the train relative to the observer shall cause a transverse contraction in the train we must expect the axles to shorten, causing the wheels to leave the rails and thus bringing the train to grief. Experience has shown that no such accidents occur.

It is not, however, merely an experimental fact that transverse changes in dimensions, due to motion, do not occur: their occurrence would be incompatible with Relativity, for Relativity would have no means of taking them into account.

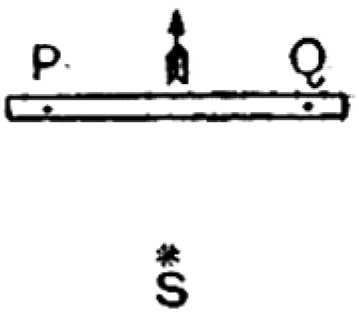

FIG. 61

We suppose the rod *PQ* to be moving away from *S* in such a manner that the distances *SP* and *SQ* remain equal as seen by *S*. Clocks fixed at *P* and at *Q* on the moving rod must appear to *S* to give similar readings (presuming that they were set together at the outset when the rod was at rest at *S*). If they gave different readings there would be no factors in terms of which to describe the difference. Gravitational fields are absent, the distances *SP* and *SQ* are equal and the relative velocities of *P* and of *Q* in respect to *S* are equal. It may be said that the *positions* of *P* and of *Q* in space are different, but it is a fundamental assumption of Relativity that position in space is indescribable and, indeed, meaningless.

But if *S* idea as to simultaneity in time at *P* and at *Q* is the same now the rod is in motion as it was when the rod was at rest it follows that both *S* and *S'* (an observer located on the rod) must form the same estimate of the distance *PQ*.

To make this clear we suppose that *S* holds his own measuring rod (at rest relative to himself) parallel to the rod, in such a way that the points *P* and *Q* slide across the divided scale of *S* rod. If *S* sees *P* slide across, say, the zero of the scale and *Q* slide across, say, the division 5 then *S'* must also see the same events, for the contacts of *P* and *Q* with the divisions on the scale are space-time coincidences and are "real" and must be seen similarly by all observers. In order to measure, by means of a scale, the distance between two points *P* and *Q*, it is necessary to observe which scale divisions touch *P* and *Q* simultaneously. We have agreed that "simultaneously" means the same to *S'* as it does to *S*, hence since *S* observes that *P* falls on *O* and *Q* on 5 simultaneously *S′* must observe the same. It remains only to show that *S′* regards *S* scale (held fixed while he slides over it) as a "true" scale. We consider, of course, the instant at which *P* and *Q* are *in contact* with the points *O* and 5. At this instant the motion of *P* relative to *O* and that of *Q* relative to 5 is "transverse" and hence not definable (as relative motion) at all. Thus *S′* regards *S* scale as "at rest" and therefore "true."

We see then that the Theory of Relativity can allow of no ambiguity as to the transverse dimensions of moving bodies, that is to say it can conceive of no observer whose estimate of the transverse dimensions of moving bodies can differ from the estimate made by other observers.

The sense in which the term transverse motion may be used

We have seen that, in respect to one particle, it is meaningless to speak of the transverse motion of another particle, and that there is no criterion available as to whether or not *P* is rotating about *O* (*P* and *O* being a pair of particles at a fixed distance apart). Similarly, in the absence of other bodies to form a frame of reference, it is meaningless to say, for instance, that the rod *OP* (Fig. 62) is rotating about *O*. Yet how can this contention be reconciled with the "fact" that if *P* rotates about *O* there is a centrifugal (gravitational) field of force along *OP* and that, for instance, a clock fixed on the rod at *O* will go more quickly than another fixed at *P*?

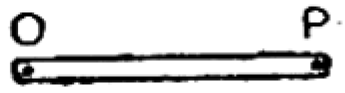

FIG. 62

The answer is that the relative going of clocks at *O* and at *P* is *not determinable*, or indeed is meaningless, except in reference to things external to *OP*. It would be natural for an observer living at *O* to define time in terms of the ticks of a clock of his own construction and to postulate that a similarly constructed clock at *P* would go at the same rate as one at *O*. No observations which he could make would ever lead this observer to doubt the legitimacy of his arbitrary ideas about time—indeed there is no sense in disputing their "legitimacy."

When—as often, in explanations of the effect of gravitation upon clocks—the statement is made that the slow going of the clock *P* would be noted by an observer sitting upon the rod at *O*, the assumption underlying the statement is that the observer at the centre *O* (where the velocity is zero) has had assigned to him a rigid 2-dimensional frame of reference (all parts of which are "at rest") in which the rotation of *OP* occurs. Thus it is assumed that the observer at *O* possesses not only his own fixed clock at *O* but also many other clocks, all synchronous with his own, in various parts of the "fixed" frame. It is then no longer meaningless to speak of the relative rates of going of clocks at *O* and at *P* on the rotating rod, for, at every point of the fixed frame, there is a fixed clock and relatively to the fixed clock near to the moving clock *P* this latter clock is in motion and suffers, therefore, the FitzGerald slowing. The fixed clock near *P* at the instant considered is synchronous with the clock at *O*, hence the *P* clock goes slowly with reference also to the *O* clock and as seen by the observer at *O*.

With this assumption, that *O* has had assigned to him a fixed, rigid frame of reference, it is no longer meaningless to speak of the transverse motion of the particle *P* relatively to the particle *O*, nor of the rotation of *OP*.

Are "ether" explanations of the Fitz-Gerald effects legitimate?

No explicit mention has been made of the Newtonian "ether," in regard to the changes in rods and clocks, deduced by consideration of the *TOX* diagrams. The method of the *TOX* diagram is, however, essentially an ether type of method, and the study of it is helped by a preliminary study of the more usual ether methods, in which the contraction of moving rods is spoken of as a direct effect of the rod's motion through the ether. These direct ether methods are legitimate and depend in no sense upon the "existence" or non-existence of a Newtonian ether, for always the postulate of a unique ether, made at the outset, is discarded later. The velocity u is taken to mean, at the outset, velocity relative to the ether, but later means the velocity of any observer relative to the event he is observing. Arguments which endeavour to adhere to the notion of an ether are evasions of Relativity, because no meaning is yet known to be attachable to the idea of motion relative to the ether; but to use the ether idea for purposes of explanation, discarding it later on in the argument, is just as legitimate as, for instance, to use the water analogy in explaining electrical phenomena, and often is very helpful.

Even those explanations which make use of flashes of lightning occurring on an embankment, and discuss the relative times of occurrence of these flashes referred, first to the embankment and then to a moving train upon it, are using, in principle, the ether idea. The tacit claim is made that correct conclusions may be reached in regard to the actual transmission of light by aid of assumptions as to its transmission relatively to a rigid identifiable railway embankment—although the actual transmission medium (if such exist) is certainly not the embankment. This seems surely to be equivalent to the assumption that, whatever may be the actual mode of transmission of light, the description of its transmission by reference to a rigid identifiable frame will lead to conclusions which will prove consistent with the observed phenomena of light's actual transmission. But such, precisely, is the assumption made by those who argue in terms of the Newtonian ether.

What is there of "reality" in the TOX diagrams?

We are now prepared to consider what there is of "reality" in the space-time diagrams *TOX* and *T'OX'* of the two observers *S* and *S'*. Each observer has his own peculiar views about length intervals and time intervals. Can we find some combination of these intervals, which shall be observed similarly by all observers?

We consider two point-events *P* and *Q* (Fig. 63) having coordinates $x_p t_p$ and $x_q t_q$ in *S* system and $x'_p t'_p$, $x'_q t'_q$ in *S'* system.

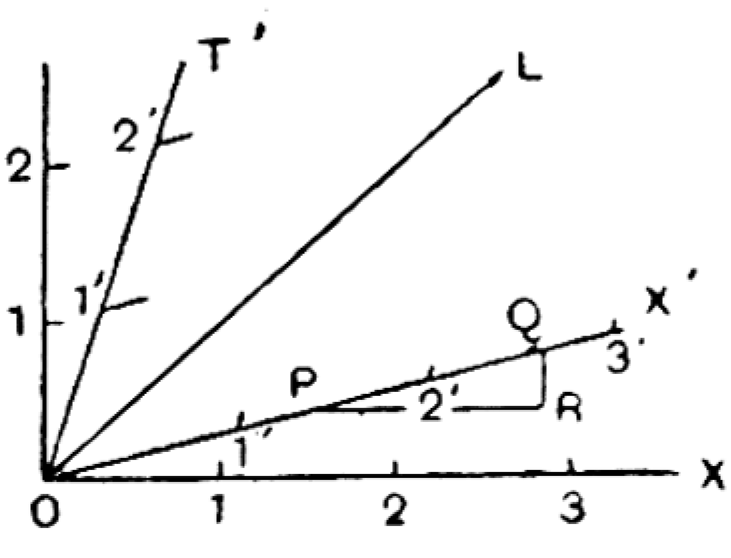

FIG. 63

We have, in S system, $t_q - t_p = QR$ paper distance and $x_q - x_p = PR$ paper distance, while in S' system,

$$t'_q - t'_p = \text{zero},$$

and

$$x'_q - x'_p = PQ.$$

"Intervals" and distances compared

Obviously the measured distance between P and Q is not the same for S as for S', for when S makes the measurement by means of clock and rod he finds P and Q separated by distance PR and time QR, or, in his own units, by (PR inches $\div$ 01 inches) units of length and by (QR inches $\div$ 01 inches) units of time. S' finds P and Q separated only in space, namely by ($PQ/01'$) units.

In an ordinary problem of mensuration in which P and Q represent, say, a pair of points on opposite sides of a pond in a field (Fig. 64), the pair of points being so located that direct measurement from P to Q is inconvenient, it is quite likely that different observers may use different systems of coordinates in measuring the distance PQ. For instance, one observer will use axes x and y parallel to PS and to SQ respectively, another will use axes parallel to PR and RQ, while a third uses axes parallel to TQ and PT. The first will find

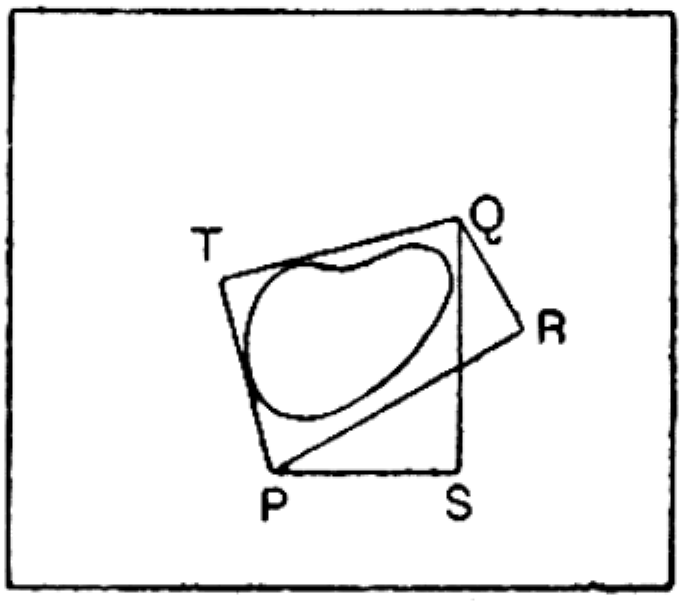

FIG. 64

$$(x_q - x_p) = PS,$$

and

$$(y_q - y_p) = QS.$$

The second will find

$$(x_q - x_p) = PR,$$

$$(y_q - y_p) = QR,$$

and the third

$$(x_q - x_p) = TQ,$$

$$(y_q - y_p) = TP.$$

Every observer, however, will get the true result for the distance D, between P and Q, if he writes

$$D^2 = (x_q - x_p)^2 + (y_q - y_p)^2.$$

This simple rule does *not* apply to measurements made by observers S and S' in our space-time field. The question is, can we find some other function of $(x_q - x_p)$, the space interval, and $(y_q - y_p)$, the time interval, which may give to all observers one and the same value?

Mathematical investigation of the geometry of Fig. 63, using various, differently oriented axes, shows (as is proved on p. 242) that every observer, whatever may be his (uniform) velocity relative to S, will find the same value for the quantity

$$-(x_q - x_p)^2 + (t_q - t_p)^2,$$

expressed, of course, in his own units of x and t. This quantity can be regarded, therefore, as expressing the "real" relation in space-time of a pair of particles P and Q. It is called the space-time interval (or simply the interval) between P and Q.

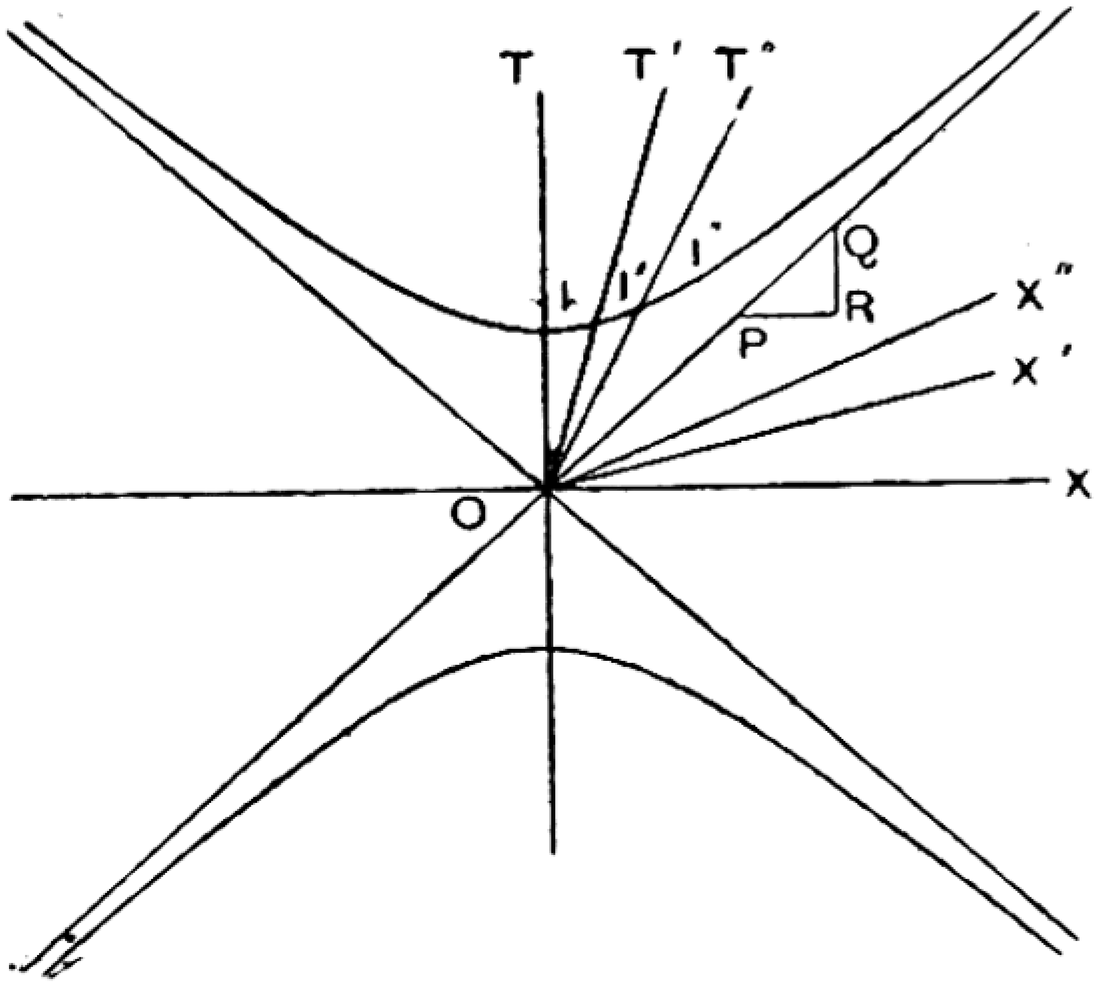

FIG. 65

The geometrical properties of the systems of axes (TOX, $T'OX'$, $T''OX''$ etc.) which have to be used by different observers are simpler than might have been expected, for it is found that the points 1, $1'$, $1''$, $1'''$ etc. on the T axes of observers S, S', S'', S''' etc., having different velocities relative to S, lie on a hyperbola, as indicated in Fig. 65.

The T axis and the X axis are always equally inclined to OL, so that the hyperbola is asymptotic to OL. This means that an observer moving with light's velocity would be employing infinitely great units of length and of time.

Some light may be thrown upon the necessity of the asymptotic form of the 1, 1′, 1″ curve by simple reasoning. This curve is the locus of the points defining the paper-units of time of the observers whose time axes are TO, $T'O$, $T''O$ etc., that is to say it is the locus of all points which are at unit "interval" from O in the space-time continuum. The "interval" between any pair of points in the diagram has the same value to every observer, hence since the points O and 1‴ are separated by unit interval to S''', they must have unit interval separation to every observer. If then the curve is asymptotic—that is to say always approaches the OL line but goes to infinity before touching it—we must understand that, to the observer whose OT axis lies along OL, a unit of time (as also a unit of length) must be of infinite paper-length.

Hyperbolic geometry

That this must be so—and that, therefore, the curves must be asymptotic—may be deduced from the fact that, *to every observer the interval separating any pair of points on OL is nil.* We consider the interval length between say P and Q in Fig. 65. To S this interval is

$$PQ = \sqrt{QR^2 - PR^2} = 0;$$

hence to *every* observer it must also be nil, including, of course, the observer whose time axis is OL. But to this observer the line OL must be the X axis as well as the T axis (because, always the X and the T axes are equally inclined to OL); hence, as both P and Q lie along OT', the triangle PQR must be such that P and R coincide, that is to say $PR = 0$ and $QR = PQ$. Hence the interval

$$PQ = \sqrt{QR^2 + PR^2} = \sqrt{QR^2} = \text{the paper-length } PQ.$$

We know also, however, that for any pair of points both on OL the interval is nil; hence $PQ = 0$.

The only way in which this statement that the interval separating P and Q is nil (i.e. zero units of time) can be made consistent with the statement that the same interval is of paper-length PQ, is by saying that the paper-length PQ is nil compared with the paper-length of one unit of time. Thus the observer S' whose axes OX' and OT' coincide with OL must use units whose paper-length is infinitely great.

The Lorentz Transformation

On pages 220 to 223 expressions have been deduced which enable us now to write down the set of equations called the "Lorentz Transformation." These are:

$$\left.\begin{aligned} x' &= \beta\,(x - ut) \\ y' &= y \\ z' &= z \\ t' &= \beta\,(t - ux) \end{aligned}\right\} \text{ where } \beta \text{ stands for } 1/\sqrt{1 - u^2}.$$

If instead of unity the symbol C be used for the velocity of light, revision of the process by which these equations have been derived shows that they take the form

$$\left.\begin{aligned} x' &= \beta\,(x - ut) \\ y' &= y, \quad z' = z \\ t' &= \beta\,(t - ux/C^2) \end{aligned}\right\} \text{ where } \beta \text{ stands for } 1/\sqrt{1 - u^2/C^2}.$$

These equations express, by the dashed letters, the values which an observer S' whose frame (x', y', z', t') is moving with velocity u in the x direction relatively to S whose frame (x, y, z, t) is "at rest," will assign to the co-ordinates of a point-event whose co-ordinates in S system are x, y, z and t.

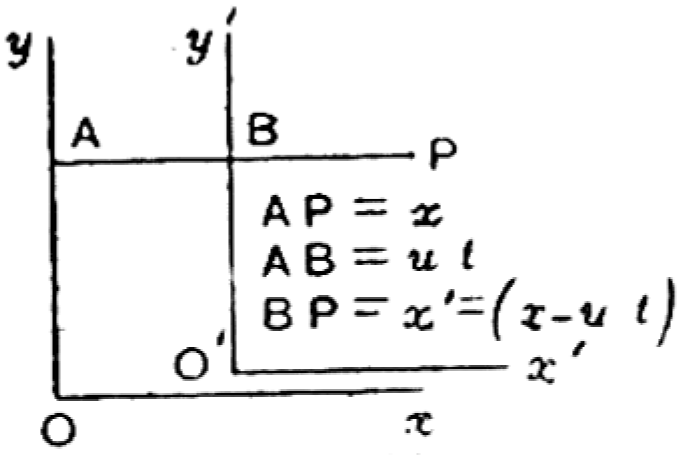

FIG. 66

In deriving the x' equation we imagine that at the time $t = t' = 0$ the frames yOx and $y'O'x'$ coincide, but that $y'O'x'$ is moving along Ox with velocity u relative to yOx. Then, after time t, it is clear that $O'y'$ will have been displaced by the distance ut to the right, and a point P distant x from Oy will be distant $(x - ut)$ from Oy'. This distance $(x - ut)$ measured by S' with his contracted scale will be called by him

$$x' = \beta (x - ut).$$

The equations $y' = y$ and $z' = z$ express the fact already discussed, that transverse changes of length cannot occur; while the t' equation has been dealt with completely on pp. 220–223.

Derivation of the true expression for relative velocity

According to Newtonian principles if the particle P in Fig. 66 is moving parallel to Ox with velocity ω relative to the S frame yOx, then its velocity relative to the S' frame $y'O'x'$ must be $\omega' = (\omega - u)$. This rule gives wrong results in the case of light for we know that when $\omega = C$, $\omega' = C$ also.

In order to find the true formula for relative velocity—a formula reducing to $\omega' = (\omega - u)$ when u is small and to $\omega' = u$ when u approaches the value C—we apply the Lorentz Transformation.

Given, $\omega = x/t$, we have to find $\omega' = x'/t'$ in which

$$x' = \beta (x - ut) \text{ and } t' = \beta (t - ux).$$

The steps are

$$\omega' = \beta (x - ut)/\beta (t - ux),$$

$$\omega' = \left(\frac{x}{t} - u\right)\bigg/\left(1 - u\frac{x}{t}\right) = (\omega - u)/(1 - u\omega).$$

Thus the true expression for relative velocity is

$$\omega' = (\omega - u)/(1 - u\omega),$$

or putting C for the velocity of light, instead of unity,

$$\omega' = (\omega - u)/(1 - u\omega/C^2).$$

An alternative method of derivation of the FitzGerald contraction

We have explained, on pp. 214–217, the derivation of the FitzGerald effects from the geometry of the TOX diagram. A more direct method will now be given, which seems perhaps to come a little nearer to making possible a "physical conception" of the phenomenon.

Three assumptions are made:

(1) The constancy of C.

(2) The unambiguity of u the relative velocity of S and S' as measured by S and S'.

(3) The non-occurrence of transverse changes of the dimensions of bodies due to uniform motion.

The argument is really an "ether" argument, since it involves the notion of "rigid" solids to act as frames of reference.

We imagine two rigid flat sheets of corrugated steel A and B (Fig. 67) of unlimited extent: they are in contact face to face and groove on groove and can slide in their grooves the one along the other with any desired velocity (u). By assumption (3) the corrugations will continue to fit properly together no matter how great the velocity u may be. Thus the corrugations may be replaced, for the sake of argument, by equally spaced parallel lines drawn upon flat sheets.

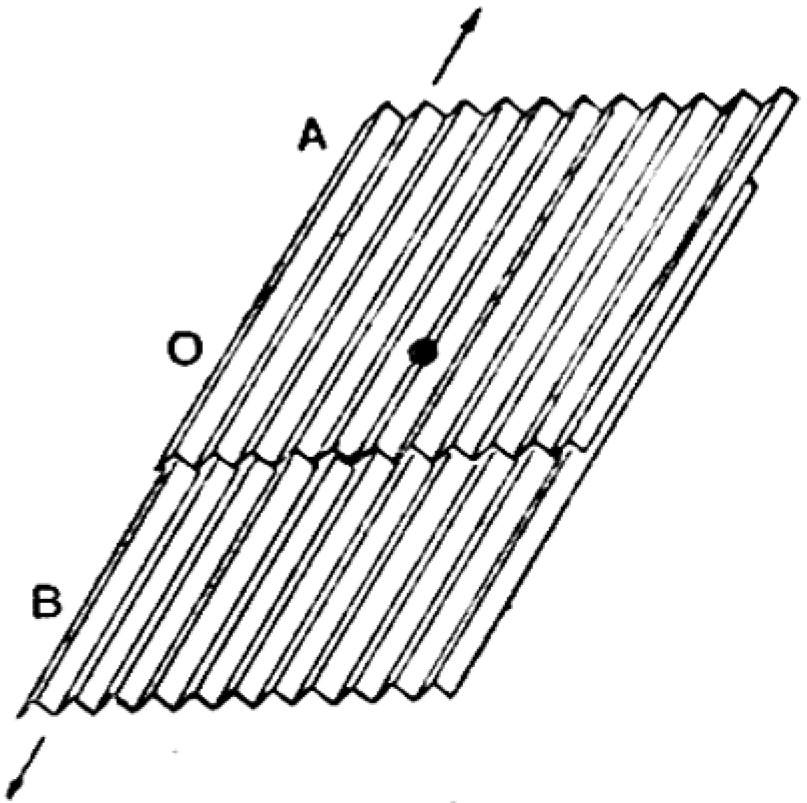

FIG. 67

While the sheets are in motion a white-hot bullet is fired at them, penetrating both sheets simultaneously, entering at O in sheet A and leaving by the coincident point O' in sheet B. We imagine an observer S at O and another S' at O', each fixed to his own sheet of steel. Both S and S' observe the white-hot bullet as a flash of light, and by assumption (1) each discovers by actual measurement, that the light, starting at the hole spreads over *his own* sheet radially in all directions with velocity C. Each finds that at successive instants of time the area of surface illuminated by the light is bounded by successive concentric circles centred on O (and O'). For instance, S, standing at O, where the flash occurred, finds, by actual measurement, that after, say, one millionth of a second as measured on his own clock the area 1 had become illuminated, after two millionths of a second the area 2 and so on (Fig. 68).

S′ obtains a diagram exactly like that obtained by *S*, but centring on *O′*. This is a direct consequence of assumption (1). Yet how absurd it seems! For when *S* is observing, say, the circle 3, *S′* sheet has shifted away so that the hole is now at *O′* (Fig. 69), and if, as we maintain, *S′* draws a series of circles precisely similar to *S* series, we arrive at the contradictory conclusions that the illuminated patch must be in two positions at once, namely, it must occupy the position of the *O* circle in Fig. 69, and, *at the same time*, that of the *O′* circle.

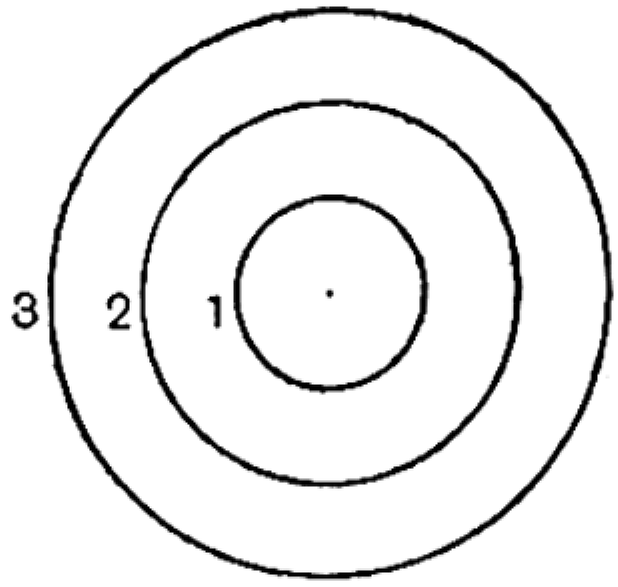

FIG. 68

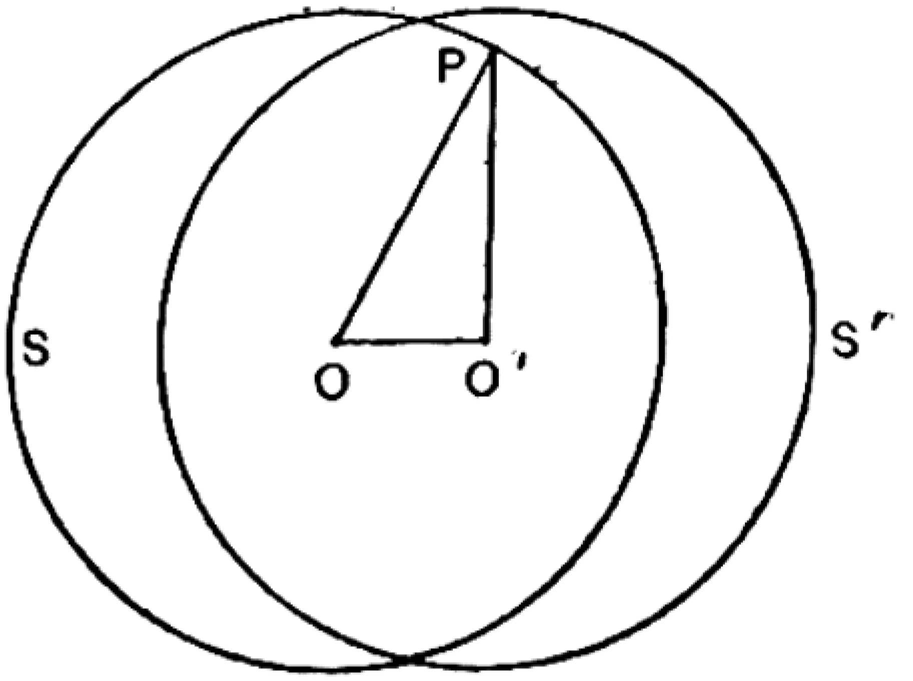

FIG. 69

The absurdity is avoidable by postulating—as Relativity does—that the expression "at the same time" is ambiguous, that is to say, that *S* and *S′* clocks have different settings on account of their relative motion.

The matter may be expressed quantitatively by reference to Fig. 69.

Let OO' be along the x axis of both *S* and *S′*—so that the corrugations represent equally spaced mesh-lines parallel to the x axis.

The relative velocity of *S′* and *S* is u, therefore *S* writes $OO' = ut$ at the time t seconds on *S* watch after *O′* left *O* (that is to say after the flash occurred).

The slowing of moving clocks

Now by assumption (3) the distance $O'P$ perpendicular to OO' being *transverse* to the motion is measured similarly by S and S'. Let both observers, therefore, measure the time taken for the light to reach the point P. S writes it t, for he knows that t seconds as measured by his own clock, have elapsed since the flash occurred. Relative to S the light has travelled the distance OP and S writes, therefore, $Ct = OP$. S' calls the time taken for the light to reach P, t', the time which *his own* clock tells him has elapsed since the flash occurred, and as we see by reference to Fig. 69 that relative to him (S') the light has travelled only the distance $O'P = \sqrt{OP^2 - OO'^2}$, where $OP = Ct$ and $OO' = ut$, we write

$$O'P = \sqrt{C^2t^2 - u^2t^2}$$

and we have to conclude that

$$\frac{t'}{t} = \frac{\sqrt{C^2t^2 - u^2t^2}}{Ct}$$

or

$$t' = t\sqrt{1 - \frac{u^2}{C^2}},$$

an equation which, when C represents unit velocity, becomes

$$t' = t\sqrt{1 - u^2}.$$

This equation means that S' watch goes slowly as seen by S, each of S' seconds being $(1/\sqrt{1 - u^2})$ times one of S seconds.

The FitzGerald contraction is deducible from the slowing.

By assumption (2) both S and S' find their mutual relative velocity to be u. Thus S writes

$$OO' = l = ut,$$

while S' writes $$OO' = l' = ut'.$$

Hence, $$(l'/l) = (t'/t)$$

or $$l' = (l/t)\,t' = (l/t)\left(t\sqrt{1 - \frac{u^2}{C^2}}\right)$$

or $$l' = l\sqrt{1 - \frac{u^2}{C^2}}.$$

This means that l', the distance from O' to O on the moving sheet B, is $(\sqrt{1 - u^2/C^2})$ times l, the distance between the *same points* on the fixed plate A. The quantity $(\sqrt{1 - u^2/C^2})$ is less than unity (since u is always, we assume, less than C), hence the *moving sheet has contracted* on account of its motion. This is the Fitzgerald contraction.

The invariant property of "interval"

We stated without proof, on p. 167, that the quantity

$$(t_q - t_p)^2 - (x_q - x_p)^2$$

was similarly measured by all observers. This may be proved very easily.

Let a number of observers S, S', S'', S''' etc. pass at a given instant $(t = t' = t'' = t''' = 0)$ through the origin O of S system along the axis Ox with uniform velocities, u', u'', u''', u'''' etc.

At time t' seconds by S' clock after passing through O, S' deduces that S clock must read $t = t'\sqrt{1 - u'^2}$.

Similarly S'' deduces that $t = t''\sqrt{1 - u''^2}$, and S''' deduces that $t = t'''\sqrt{1 - u'''^2}$ and so on. Hence:

$$t = t'\sqrt{1 - u'^2} = t''\sqrt{1 - u''^2} = t'''\sqrt{1 - u'''^2}$$

and so on, or,

$$\sqrt{t'^2 - u'^2 t'^2} = \sqrt{t''^2 - u''^2 t''^2} = \sqrt{t'''^2 - u'''^2 t'''^2}$$

and so on.

But since in S' system $u't' = x'$ and so on we have

$$(t'^2 - x'^2) = (t''^2 - x''^2) = \text{etc.}$$

This quantity is an "invariant" because all observers (in uniform relative motion) obtain the same numerical result when they measure it.

For the general case in which the relative motion may be in any direction (not, as we have assumed, merely along the x axis) we use the fact that if a particle moves from O in time t to a point P whose co-ordinates are x, y, z then

$$OP^2 = x^2 + y^2 + z^2.$$

Then if the relative velocity u is along OP we have

$$OP^2 = u^2t^2 = x^2 + y^2 + z^2.$$

Thus the "invariant"

$$(t^2 - x^2) = (t^2 - u^2t^2)$$

may be written for the general case:

$$\{t^2 - (x^2 + y^2 + z^2)\}.$$

It should be noted that, in this expression, t, x, y and z stand respectively for the time taken and for the distances traversed (at constant velocity u in the OP direction) during that time. For cases in which the velocity is changing (that is to say the motion of the particle considered is accelerated) it is permissible to regard u as constant for a short time dt during which the distance moved through by the particle is

$$udt = \sqrt{dx^2 + dy^2 + dz^2}.$$

Hence the form taken by the invariant in the more general case in which acceleration is not excluded is

$$dt^2 - (dx^2 + dy^2 + dz^2) = ds^2.$$

The invariant as thus stated refers to the case in which the system of units chosen makes $C = 1$. If we go back and write $(\sqrt{1 - u^2/C^2})$ instead of $(\sqrt{1 - u^2})$ we shall find that, instead of dt^2, the quantity C^2dt^2 occurs in the expression for the invariant, which then reads:

$$C^2dt^2 - (dx^2 + dy^2 + dz^2).$$

It is interesting and instructive to reverse the process which has just been followed.

We suppose that a light ray travels from a point P to a point Q in S frame in time dt. Then

$$Cdt = \text{distance } PQ.$$

But, by geometry,

$$PQ = \sqrt{dx^2 + dy^2 + dz^2}$$

where dx, dy and dz are the space components separating P and Q. Hence

$$Cdt = \sqrt{dx^2 + dy^2 + dz^2},$$

or

$$C^2dt^2 - (dx^2 + dy^2 + dz^2) = 0.$$

This is true for all observers, not for S only; all using the same value for C, while they measure by means of their own instruments, and in their own units, the space and time intervals. Thus the quantity on the left of the above equation is invariant. This quantity may be multiplied by any desired constant (e.g. by $1/C^2$) and the square root of the result may then be taken without destroying the invariance. Thus

$$\sqrt{\frac{C^2dt^2 - (dx^2 + dy^2 + dz^2)}{C^2}} \text{ is invariant,}$$

that is

$$\sqrt{dt^2 - \frac{dt^2}{C^2}\left(\frac{dx^2 + dy^2 + dz^2}{dt^2}\right)} \text{ is invariant,}$$

or

$$dt\sqrt{1 - \frac{1}{C^2}\left(\frac{u^2dt^2}{dt^2}\right)} \text{ is invariant,}$$

i.e.

$$dt\sqrt{1 - u^2/C^2} \text{ is invariant.}$$

The symbol dt represents the time interval during which any space path $(dx + dy + dz)$ is being traversed *by anything*. If light is the thing then C^2dt^2 is numerically equal to

$$(dx^2 + dy^2 + dz^2).$$

If a particle having velocity u is the thing travelling then u^2dt^2 is numerically equal to $(dx^2 + dy^2 + dz^2)$.

Deduction from the Lorentz Transformation, of the effect of gravitation upon rods and clocks

By the Principle of Equivalence any gravitational field may be reproduced, in a small region, by an acceleration of the frame of reference. The equations of the Lorentz Transformation, as for instance

$$t = \beta\,(t' - ux'/C^2),$$

refer to the effect on rods and clocks, of a constant velocity u. In order to apply them to the phenomena of gravitation we must adapt them to *changing velocities*. Moreover since both uniform velocities and gravitational acceleration have their effects upon rods and clocks, it is desirable, in order to separate out the gravitational effect, so to adapt the equations that while a (du/dt) appears, u disappears.

We take the Lorentz equation just given and understand that S' frame, with a clock distant x' from O' (in S' units), begins to move from rest with acceleration (du/dt) relative to S' in the direction Ox' and that, at the instant under consideration, O and O' coincide and $t = t' = 0$. We postulate that du/C is negligible compared with unity, so that to take its actual value into account in the equation can have no greater effect than might result from a mere error of observation. Can we put the equation into a form in which the effect of velocity is negligible while that of acceleration is expressed?

Acceleration is u/t, or, if changing, is du/dt. We put the equation, therefore, into the form

$$dt = \beta\,(dt' + du\,x'/C^2) = \frac{(dt' + du\,x'/C^2)}{\sqrt{1 - du^2/C^2}},$$

that is

$$dt = \frac{dt'\left(1 + \dfrac{du}{dt'} \cdot \dfrac{x'}{C^2}\right)}{\sqrt{1 - du^2/C^2}},$$

and we now suppose that S' frame, and his clock, fixed at distance x' from his origin O' which, at the instant considered, coincides with O have had their velocity raised from nought to du in time dt'.

Then since u/C is negligible we neglect the denominator and write

$$dt = dt'\left(1 + \frac{du}{C^2}\frac{x'}{dt'}\right) = dt'\,(1 + a\,x'/C^2),$$

where a is the acceleration of S' frame and clock, relatively to S frame.

The reason we can neglect u/C in the denominator yet cannot neglect du/C^2 in the numerator (although the value of du/C^2 is infinitely small) is this, that in the numerator

du/C^2 is multiplied by x'/dt' and although du/C^2 is infinitely small, x'/dt' may be infinitely great (since x' has any desired finite value). The product (nought × infinity) may have *any* value, and to find what the value is in our particular case we must work out the actual value of the expression

$$(du/C^2)\,(x'/dt').$$

We note that $(1 + a\,x'/C^2)$ is greater than unity if a is positive —which means when u is positive and, as we have defined distances to the right from O to x as positive, this means when S' velocity away from S is increasing. If now we apply the formula to the watch in the pocket of, say, a man falling to earth from an aeroplane, giving both to him (S') and to an observer S, sitting on *terra firma*, frames of reference whose origin—*at the instant considered*—is the earth's centre, we see first that x' in the equation means r the radial distance of the airman from earth's centre, and next that a must be called *negative* because the airman and his frame are *approaching* the ground-man and *his* frame.

Thus the equation is

$$dt = dt'\,(1 - a\,x'/C^2)$$

where a is the numerical value of the airman's acceleration and x' is his distance from earth's centre; dt' being the airman's estimate of the time interval between, say, his leaving the aeroplane and the explosion of the petrol tank, while dt is the ground-man's estimate of the time interval between *the same events*. The formula says that dt is less than dt', which means that the ground-man's clock (in the earth's field of gravitation) goes *more slowly* than the airman's clock—which, falling freely, is freed from gravitation.

Gravitational potential

The equation $dt = dt'\,(1 + a\,x'/C^2)$
is often written $dt = dt'\,(1 + \phi/C^2)$
where ϕ (= ax) is called the *potential of the gravitational* field at the point distant x' from the origin.

We regard S frame as having its origin at, say, the earth's centre, and assume that at the instant considered (that is to say the instant at which *dt* occurs or begins) S' frame has its origin also at earth's centre. x' the distance of S' clock from O' is thus equal to the radial distance of this clock from earth's centre.

The work done in transporting a particle of unit mass from earth's centre to a distance x' is ax'*; this work is called the potential at x' and is represented often by the symbol ϕ. Thus our equation gives the slowing effect upon a clock, of a gravitational field of known potential ϕ.

Behaviour of measuring rods in a Gravitational field

Since, by the Principle of Equivalence, to put an observer at a spot in a gravitational field is equivalent to putting him (in a region where no permanent gravitational field exists) into an accelerated lift, that is to say to giving him a certain sort of motion, and since, by the principle of the relativity of motion, such a man may regard himself, legitimately, as at rest, it follows that such a man must find for the velocity of light that same numerical value C which an observer at rest in a region free from gravitation would find. He has a right, therefore, to use his rod both as a clock and as a rod, that is to say as a unit of length and also (in the manner explained on p. 204) as a unit of time.

Thus, for the region of that observer, it is legitimate to draw

---

* Let x be the radius of earth $=R$. Then the potential at this radius (defined as the work—in this case negative—done in bringing unit mass from an infinite distance—or practically from any distance so great that the earth's field is undetectable there—to the region concerned) is equal to

$$W=\int_{R=\infty}^{R=R} FdR$$

and F, the force on unit mass, is, by Newton's law, equal to KM/R^2 where M is the earth's mass and K is Newton's gravitational constant.

Thus $$\phi=W=\int FdR=KM\int_{\infty}^{R}\frac{1}{R^2}dR=KM/R.$$

But F, the force on unit mass $=m\times a=1\times a=a$ and we have seen that $F=KM/R^2$, hence $a=KM/R^2$, and since $x'=R$, therefore

$$ax'=KM/R=\phi.$$

a small *T′OX′* diagram, and to use all the arguments already given in reference to such a diagram, in order to determine the relative going of clocks and the "changes" occurring in the lengths of rods, and we conclude that in this small gravitational region whatever factor expresses the lengthening of the "time" between two clock ticks must also enter, as before, into the expression of the contraction suffered by rods.

We have seen that the effect on clocks is given by

$$dt = dt' \, (1 + a\,x'/C^2) = dt' \, (1 + \phi/C^2),$$

hence we conclude that the equation showing the effect of a gravitational field upon rods must be

$$dl = dl' \, (1 + \phi/C^2)$$

in which ϕ usually is negative (see p. 247), and so dl', the greater value, refers to the observer who is *in* the field, whose *rod* suffers contraction.

CHAPTER XII

Curvilinear motion

For simplicity we have been considering motion along a straight line, that is to say motion in space of only one dimension or in 2-dimensional space-time. For motion in 3-dimensional space or 4-dimensional space-time the following equation, analogous in form to that for two dimensions, is found to hold (see p. 243):

$$s^2 = -(z_p - z_q)^2 - (y_p - y_q)^2 - (x_p - x_q)^2 + C^2(t_p - t_q)^2.$$

We shall use a system of units in which $C = 1$.

What does this equation mean?

I am in a cubical room (Fig. 70) and I observe two particles P and Q. Using the corner lines Ox, Oy and Oz as space axes, I locate P by measuring, $PM = x_p$, $PN = MW = y_p$ and $OW = z_p$ and I locate Q by measuring, $x_q =$ zero, $Oy = y_q$ and $Oz = z_q$.

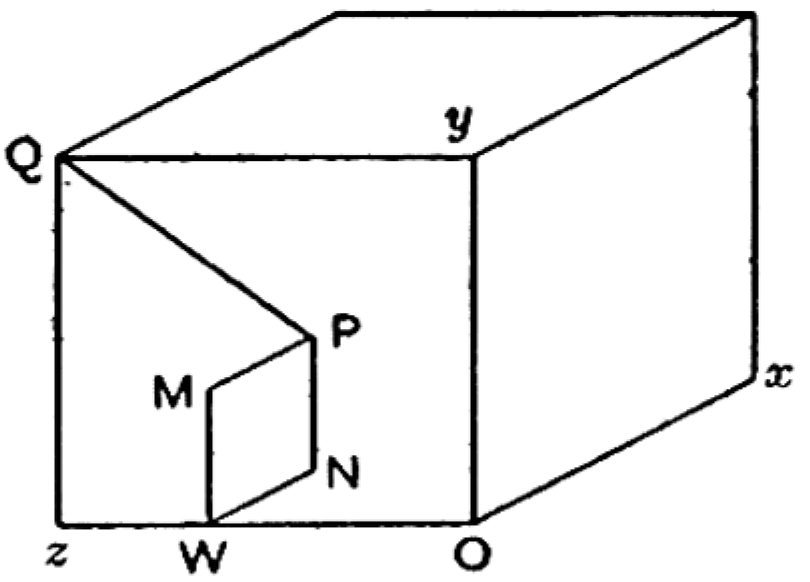

FIG. 70

The two particles may have made their appearance simultaneously according to my clock, in which case $(t_p - t_q) =$ zero, or P may have appeared at P some t seconds later than Q appeared, in which case I write $(t_p - t_q) = t$. These measurements enable me to calculate s^2. Thus

$$s^2 = (PM - O)^2 - (PN - Oy)^2 - (OW - Oz)^2 + t^2.$$

As C is to be unity we may measure t in seconds and PM PN, Oy, OW and Oz will then have to be measured in units of length each equal to 186,000 miles (or 3×10^{10} centimetres). Alternatively I may measure the distances in, say, metres in which case the unit of time in which t is to be measured must be the time taken for light to travel one metre, namely $(1/3 \times 10^8)$ of a second.

It is interesting here to note how very unimportant in the "real" world are ordinary space intervals as compared with ordinary time intervals. We choose a yard as a convenient practical unit of length and a second as a convenient practical unit of time, as though they were in some sort of practical sense quantities of like order, yet we have just seen that in the world of Relativity one second is worth some 300 million yards! Thus phenomena are very long drawn out in time, or, to put it differently, the world-lines of all particles any individual is likely to observe are nearly parallel to that individual's axis of time, or, in the *TOX* diagram, the angle *TOT′* is very small for almost all material particles (except, say, isolated electrons whose velocity relative to an observer does sometimes approach that of light).

Path of a particle in space-time

Inspection of space-time diagrams discloses a number of interesting facts about our world. For instance, we can obtain a definite answer to the question, by what path does a particle move from *P* to a "neighbouring" point *Q*? At first sight the question seems to admit of an infinite number of "possible" answers, but consideration will show that, in fact, the particle can travel only by one path, namely, along the straight line *PQ*, and that this is the longest possible path.

P and *Q* are supposed to be "neighbouring" points, by which it is meant that the space-time region between them may be regarded as approximately flat. It is realized, of course, that "neighbouring" points may be what we should call far apart spatially: in regions spatially remote from gravitating matter they may even be far apart in time also.

It must surely be granted that if a particle moves from *P* to *Q* it must follow some *definable* path and that this must be the path which it "really" follows. As the path is "real" it must appear similarly to all observers, its description being understood similarly by all observers. It is useless to describe the path's space length or its shape, because length and shape are relative and ambiguous properties, nor is it any use saying that the path has "interval length *S*," since there are an unlimited number of paths of equal interval length. Interval length is defined in 2-dimensional space-time as

$$S = \sqrt{(t_2 - t_1)^2 - (x_2 - x_1)^2},$$

and it is clear that such paths as, say, *PRQ* and *PWQ* (Fig. 71) and a large number of other paths might be chosen, all having the stated value of the interval length. The interval separating *P* and *Q* (by the path *PRQ*) is the interval *PR* plus the interval *RQ*, while by the path *PWQ* it is the interval *QW* plus the interval *PW*. With a little trouble many other paths could be found all having a total interval length equal to that of *PRQ*.

The suggestion that the particle must follow the path of shortest possible interval length is no less vain, because there are an unlimited number of such paths—paths, namely, whose interval length is nil. For instance the paths *PRQ* and *PABCDEFGQ* (Fig. 72) both have zero interval length, for they are built up of links parallel to the paths (*OL* and *OL′*) of light rays, and, obviously, for any pair of points on the line *OL* or on the line *OL′*, we have $dx = dt$ and so $s^2 = dt^2 - dx^2 = 0$.

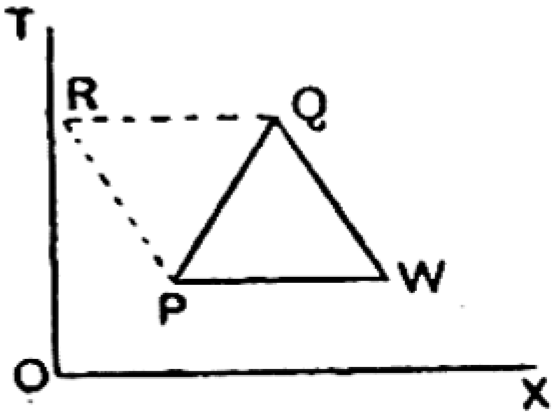

FIG. 71

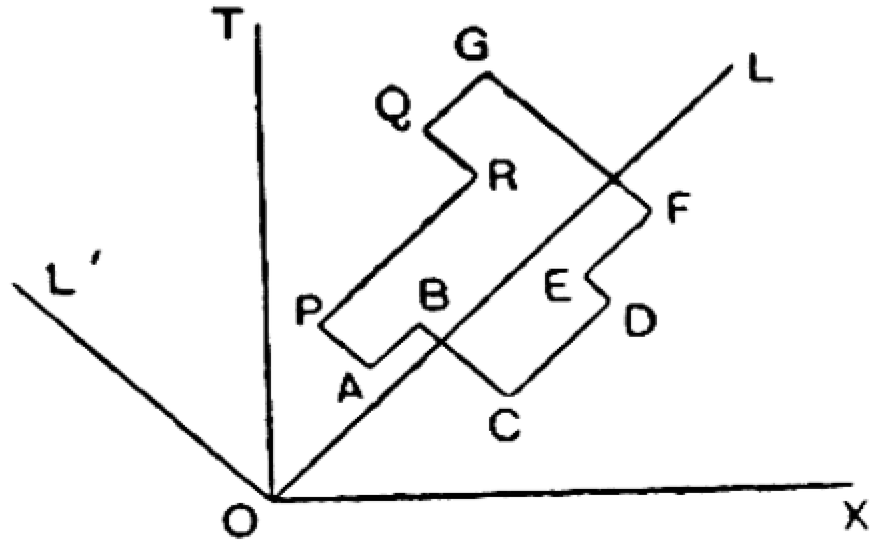

FIG. 72

In fact we soon reach the conclusion that the only *unambiguously definable* space-time path and therefore the *real* path from *P* to *Q* is *the line PQ*.

The interval *PQ* is similarly measured by all observers and we can demonstrate easily that, from every observer's point of view, it is unique, being the maximum possible interval.

To *all* observers the interval *PQ* has one and the same measured numerical value. We choose that particular observer whose time axis passes through both *P* and *Q* (Fig. 73), and for convenience we use rectangular co-ordinates. Any route, such as PWQ from P to Q, is made up of intervals such as from P to W and from W to Q.

In Fig. 73 we have:

$$(\text{Interval } QW)^2 = (QN)^2 - (NW)^2$$

because $$S^2 = (t_q - t_\omega)^2 - (x_q - x_\omega)^2,$$

hence $$[QW] < [QN],$$

the square brackets signifying "interval length."

Similarly, $$[WP]^2 = (PN)^2 - (NW)^2,$$

and $$[WP] < [PN].$$

Hence $$[QWP] < [QN] + [PN] \text{ i.e.} < [QP].$$

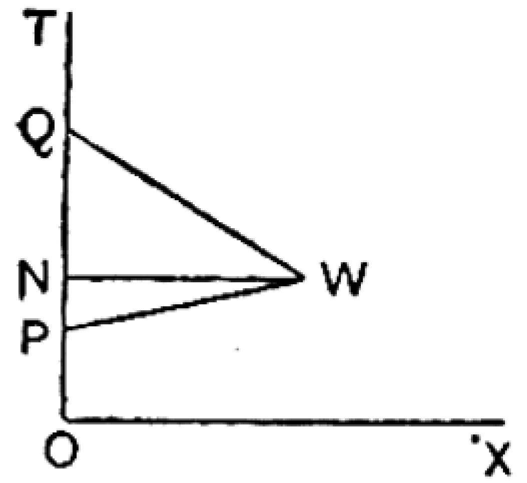

FIG. 73

By such reasoning we see that $[QP]$ is the maximum possible interval path from Q to P as seen by the chosen observer. But intervals are measured similarly by all observers, hence the square bracket equations just given are true for every possible observer and so is the result.

The path PQ is thus seen to be the only path capable of unambiguous description, and, being the maximum possible, it is unique.

To say that the Special Theory of Relativity deals only with uniform motion, is to say that it deals only with the description of the motion of particles and rays whose *longest* possible interval paths from point to point in space-time are *straight lines*.

We have described such motion by reference to geometrical diagrams to which Euclidean rules are applicable, but since in Euclidean geometry a "straight" line is the *shortest* path from point to point while in the geometry of the space-time continuum, as dealt with by the Special Theory, the straight line is conveniently used to represent the *longest* path, it is incorrect to say that the geometry of this sort of space-time is Euclidean, *Semi-Euclidean or "hyperbolic geometry"*

and the term "semi-Euclidean" is often applied to it, or, as in this sort of continuum (space-time free from hummocks) the locus of all points which are at unit distance from the origin is a hyperbola (see p. 235), the term "hyperbolic geometry" is sometimes applied.

Return to the General Theory

We return now to the General Theory, which deals not only with uniform motions but also with accelerated motions and with permanent gravitational fields. If uniform motion in space-time is indicated by a straight line then accelerated motion requires a bent line. We saw in Chap. VI that small portions of any continuous curve may be dealt with as though they are straight lines. The principle involved is familiar, for on it depends the common acceptance, as consistent with each other, of the two definitions of the tangent to a curve, namely the definition that says the tangent touches the curve without cutting it and the other that says the tangent is the straight line which joins two adjacent consecutive points on the curve. We are thus accustomed to the idea that two points very close together upon a continuous curve may be joined by a straight line which coincides with (since it does not cut) the curve.

The extension of the same principle to continua of four dimensions leads to the conclusion that small regions may be explored by Euclidean methods. In particular we can say that it will be allowable to divide up any world-line, whatever may be its shape, into small rectilinear elements (links) *ds* defined by the equation

$$ds^2 = dt^2 - dx^2 + dy^2 + dz^2.$$

We have shown already that, for uniform motion, interval as defined by such an equation is a thing which gives numerically similar measure results to all similarly equipped observers, hence if, for accelerated motion, that is to say for space-time diagrams in which the paths are not straight, an equation of similar form is to be used we must cut up the space-time continuum into very small 4-dimensional "blocks" (whatever this may mean), blocks so small that the pieces of world-lines embedded in them may be regarded as straight lines. Then, within the limits of each block, we may apply the equation just given for ds^2, and may argue in terms of *TOX* diagrams: diagrams which, in large regions, are applicable only to unaccelerated motion.

What has become of the *g*'s? This point was dealt with on p. 174. It is simply that, in considering any one of the small "blocks," we postulate an observer located in that block with units so chosen that, to him, the *g*'s for his immediate region are all unity. Such an observer uses such units provided he has been equipped with a rod and a clock calibrated in the usual way—that is to say by being brought to rest and standardized by comparison with the standards of that observer whom we have chosen to define as "at rest"—and he may be quite arbitrarily chosen. We know that such an observer, placed anywhere—no matter whether gravitational fields be there or not—and in any sort of motion, must find, by measurement in his own immediate region, that the velocity of light is unity—in other words that his own local space-time is free from hummocks (for in hummocked space—permanent gravitational fields—the velocity of light is less than unity), and that the *g*'s are all either zero or unity in accordance with the simple ds^2 equation just given.

Hence, in order to find what there is of "reality" about a world-line, which, to some particular observer and in his chosen mesh-system, appears as a certain bent line, all we need to do is to cut up this line into little links of interval length ds and say that to every possible observer the value of each ds must be the same. If then the total interval length along the whole world-line be defined as $\int ds$, the sum of all the links, this total interval length will be observed similarly by all observers: it will represent the "real" space-time separation of the beginning and end of the bent world-line.

We have seen that the straight line representing any ds link is the longest possible interval path along the link, and is the path which a particle must take if it travels from one end-point of the link to the other. If then it ever happens that we know from experiment what, in fact, is the space-time path of any thing (particle or piece of matter or ray), we have a right to conclude that each step in the path, and therefore the whole path, is the longest possible space-time interval path that thing could take.

The first law of motion of Relativity

Hence the first law of motion of Relativity is that every free particle moves so as to take the longest possible (interval) path from point to point.

We do happen to know the free paths of the pieces of matter called planets; the earth making for instance an almost circular path around the sun, a path which, plotted in 2-dimensional space, and time, with the sun as origin, is a spiral—like a cork-screw—with axis parallel to OT. This spiral path is, therefore, the longest possible space-time path which the earth could follow, that is to say the effect of the sun is so to distort space-time that the spiral path (and not a straight path) is the longest possible.

The spiral paths of planets

This fact, that the planets' space-time paths have to be regarded by us as spirals, and that, therefore, in the world around us the longest possible paths from point to point are of spiral form, while the shortest possible paths (being, as we have seen, zero) are infinitely shorter than the Euclidean shortest (straight line) paths, tells us plainly that the geometry of the real world of phenomena is certainly not Euclidean.

Thus the Relativist is compelled to describe events in terms of non-Euclidean geometry.

The geometry of the practical world is non-Euclidean

Not only is space-time geometry non-Euclidean, it is a geometry which changes from point to point in space-time. It might conceivably change in such an erratic way as to make of phenomena an indescribable chaos. It happens, however, that the geometry of our world remains practically unchanged in regions whose space-position is unchanged and that such regions are practically definable and identifiable.

This world (universe) is so constructed that the geometrical character of regions in space-time, defined by what we generally identify as "given regions in space," is fairly constant, and descriptions of phenomena are possible whose usefulness has a very long life. Moreover the changes in space-time geometry which do occur are generally such as can be predicted.

Thus, for instance, the geometry of the space-time in that region of space defined by the paths of rays of light reaching earth from certain identifiable stars is known to change in a definable way when the gravitational field of the sun lies in the paths of the rays.

Einstein predicted, and the eclipse experiments of May 1919 demonstrated, that the geometry of the region concerned would change in such a way that the paths of rays of light coming from certain stars would appear to us, on earth, to be bent in space.

Apparent reduction of light's velocity in a gravitational field

It must be remembered that, in spite of motion—whether accelerated or not—and of gravitational fields, every observer equipped with proper instruments must find for light's velocity, by actual measurement locally, the same value ($C = 1$). An observer needs only one sort of instrument, namely a rod, for this as explained on p. 204 may be used also as a clock; and any rod will do, no matter what may be its "length," provided that all observers carry *similar* rods, that is to say rods which when brought all to rest relatively to one another and superposed are found equal.

We have now to show that according to the *estimate* of an outside observer looking in from afar into a distant gravitational field, the velocity of light in that field will be less than unity.

The outside observer makes an estimate, *not* a measurement—the real *measured* velocity of light must be—according to Relativity—always and everywhere *unity*. This outside observer's estimate is useful, however, because by its aid the observer can calculate what, to him, appears as the *refraction* of light in a gravitational field. He sees light as bent in passing through a distant field, but it is not "really" bent. The bend expresses the reduction of light's velocity in the field, but this reduction is not *real*.

The deduction of the rod-contraction and clock-slowing due to gravitational fields, as given on pp. 240 *et seq.*, showed that these effects involve nothing which is not contained already in the Lorentz Transformation; the two factors, clock-setting and FitzGerald effect, accounting, between them, for all the peculiarities of the Relativist's world.

This raises the questions: 1st, why clock-settings, as such, are needed in the case of uniform motion?—the suggestion being that an increase in the contraction and slowing effects might be postulated instead, and 2nd, why setting adjustments should not be made in the case of accelerated or gravitational systems? The answer to the first question is that (as trial will readily show) rod-contraction plus clock-slowing, acting alone, would accentuate the excess above unity of the measured velocity of a light ray towards whose source the frame of the observer is moving, although, to *some extent*, they would compensate the *defect* in velocity due to a retreating frame. Thus if contraction and slowing effects alone were to be made to bring light's measured velocity always to unity, then the amounts of these effects would have to be chosen by S' to suit the direction relative to him (S') of the ray being examined. Thus S would see changes in S' rod and clock which no longer were determined by the relative motion and position of S and S' but which were determined (quite erratically as seen by S) by the direction of the particular light ray upon which S' might happen to be putting his attention!

By *setting* adjustment, however, simple trial (along the lines of the example given on pp. 220 *et seq.*) shows that in all cases of uniform motion the setting (ux') determined solely by the relation of S and S' and their frames provides the exact compensation required.

The answer to the second question, why clock-setting is not postulated in cases of accelerated systems, is that, since the setting—given by the term ux' in $(t' + ux')$—of a *given* clock for which x' is fixed, in S' system, depends upon u, therefore in cases in which S' system is accelerated and u is continuously changing the setting of that given clock would also be continuously changing.

But to say that a clock's *setting* is continuously changing (e.g. becoming more and more backward) is a clumsy way of saying that the clock is going slowly. Thus in accelerated systems it is solely the *going* of clocks which can be specified. The gravitational slowing (equivalent to a continuous re-setting) adds itself to the FitzGerald slowing (or subtracts itself, as the case may be) and the one effect cannot be separated out from the other.

We have now to consider the difficulty resulting from the fact that, although every gravitational field is equivalent locally to an accelerated frame, permanent gravitational fields exist in respect of observers whose frames possess no relative velocity. For instance, S' is on the earth's surface and S is somewhere in space right outside earth's field, and fixed relatively to S'. S sees the contraction of rods and slowing of clocks due to S' field. What does S conclude as to the value which S' must find for C? A shortened yard and a lengthened second—obviously S' must find C *greater than unity*.

The similar difficulty which threatened to arise in cases of uniform motion, on account of the FitzGerald contraction and slowing, was avoided by the postulation of various settings which compensated exactly the excess or defect (above or below unity) of the value found for C by S'. In this case S always knew the quantity u representing S' velocity relative to S, and on this he based his estimate of the setting which S' clocks must have—assuming, as Relativity demanded, that S' must be made to find $C = 1$. If S' velocity u was away from S along the direction Ox, it was to be called positive and then S' clocks had backward settings increasing with x'; while, if u was towards S, the settings of S' clocks were to be forward in a corresponding manner.

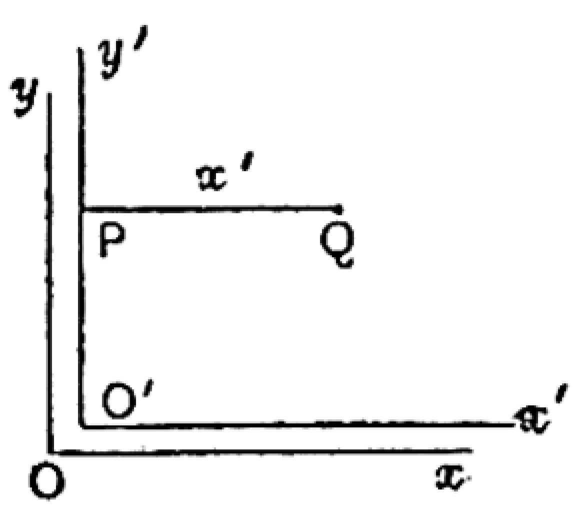

FIG. 74

But now that S and S' are a fixed distance apart u is nil and S cannot know what sort of settings to attribute to S' clocks. S merely sees that S' yard and second are respectively short and long and he has to conclude that *if indeed light's velocity is still unity in S' gravitational field*, then S', on measuring it, must find it greater than unity. Then, as Relativity forbids the possibility of S' finding C to have any value except unity, S is compelled to postulate that the light, in S' field, has a velocity less than unity.

There is no contradiction in this. S does not say that the light in S' field has at the same time two different "real" velocities. Relativity says that the real and only real velocity of light is always and everywhere unity, and it defines "real" as meaning actually measured. The observer S does not actually measure light's velocity in S' field, he merely looks on from afar and reasons—according to Newtonian methods of reasoning—about the velocity which light must have in S' field. S' on the other hand does actually measure the velocity with rod and clock; he finds its real value, and Relativity demands that this shall be unity.

Behaviour of light in a permanent gravitational field

Those gravitational fields which are equivalent over a large region to accelerated frames are to be distinguished from "permanent" fields—such as those of earth and sun—which can be treated only in small local regions as equivalent to accelerated frames.

Thus the "lines of force" of a permanent field are not everywhere parallel: being, for instance, in the case of a planet, radial. This means that the intensity of the field changes with the distance from the centre of the gravitating mass which it defines (or, which "causes it"—according to one's point of view).

The velocity of light suffers reduction in a gravitational field, and the reduction increases when the intensity of the field increases; consequently when a distant observer watches a ray approaching a gravitating mass radially he must see not only that the ray's velocity is always less than unity, but also that it is a *decreasing* velocity. To say this is equivalent to saying that the ray is repelled by the gravitating mass.

The same would be true of a bullet approaching earth with light's velocity: the bullet would suffer repulsion as seen by a distant observer.

Refraction of light in a gravitational field

It is usual to explain the refraction of light when it passes, say, from air into glass, in terms of the reduction in velocity suffered by the light on entering the glass. In a similar way precisely, the bending of light when it enters a gravitational field may be explained in terms of the reduction in the light's velocity, and as the refractive index of a medium is defined in terms of the reduction in velocity suffered by light in that medium, it is usual to speak of the refractive index of a gravitational field of known intensity. Professor Eddington gives the value 1·00000424 for the refractive index of the sun's gravitational field at the sun's surface*.

---

* *Space, Time and Gravitation*, A. S. Eddington, M.A., M.Sc., F.R.S. Camb. Univ. Press, 1920.

CHAPTER XIII

Matter

RELATIVITY deals only with "reality" as defined by the universal observer. Reality as thus defined cannot be substantiated for anything which cannot be measured. The characteristic of "matter" is that its boundaries can be measured but not its "insides." Relativity says nothing about the insides of matter; it recognizes regions in space-time characterized by hummocks (permanent gravitational fields), and the equations telling what will happen when one begins to measure up *ds* triangles in such regions show that they invariably possess an *impenetrable nucleus.*

Einstein's equation for ds^2 in a gravitational field is

$$ds^2 = -(1/\gamma)\,dr^2 - r^2\,d\theta^2 + \gamma\,dt^2$$

in which $\gamma = (1 - 2m/r)$.

The equation in this form applies to 2-dimensional space, and time (3-dimensional space-time), in which a particle makes its orbit—as a planet moving round the sun—the actual space-time path of the particle being a spiral with axis Ot.

Fig. 75 indicates the meaning of r, θ and t. The position of a particle at P in the equatorial plane of the gravitating sphere M is defined by its distance r from the centre O of the mass, and the angular distance POA (in radians) of the line OP from a fixed line OA.

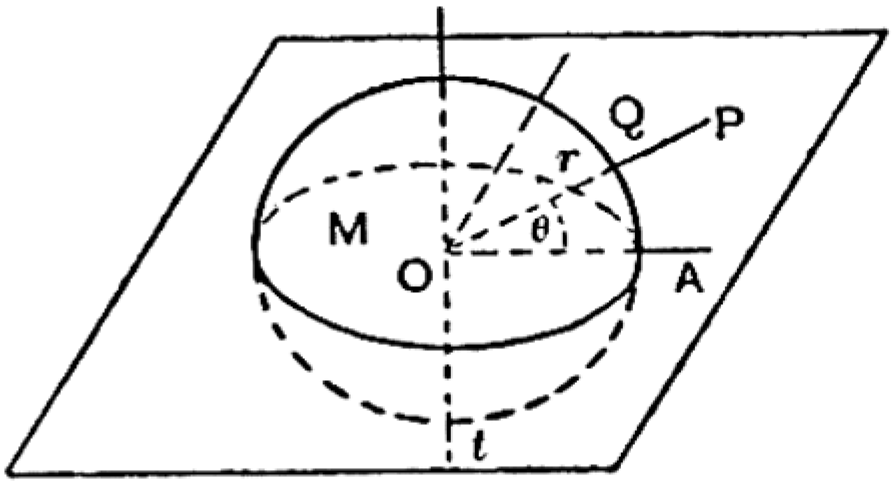

FIG. 75

We consider the interval between two points P and Q on the radius OP. The "triangle" of which PQ is the hypotenuse (*ds*) is simply the line PQ, just as, for instance, in the 3-dimensional space of Fig. 76 (using a rectangular mesh-system) PQ is the triangle of which itself is the hypotenuse. The line RP (Fig. 76) is the hypotenuse *ds* of the *ds* "triangle" whose other sides are PN ($= dx$), NM ($= dt$) and MR ($= dy$). In the same figure (Fig. 76) we may say that PQ is the *ds* triangle in which

$ds = PQ$, $dx = 0$, $dy = PQ$ and $dt = 0$ and for which the equation

$$ds^2 = dx^2 + dy^2 + dt^2$$

is

$$ds^2 = O + PQ^2 + O$$

or

$$ds = PQ.$$

Referring now to the ds "triangle" PQ of Fig. 75 whose equation is

$$ds^2 = -(1/\gamma)\,dr^2 - r^2 d\theta^2 + \gamma\,dt^2$$

we have

$$ds = PQ,\ d\theta = 0, \text{ and } dt = 0$$

whence the distance PQ *as measured in rod-lengths* $= ds$ and

$$ds^2 = (1/\gamma)\,dr^2.$$

The change in sign of γ is made for convenience.

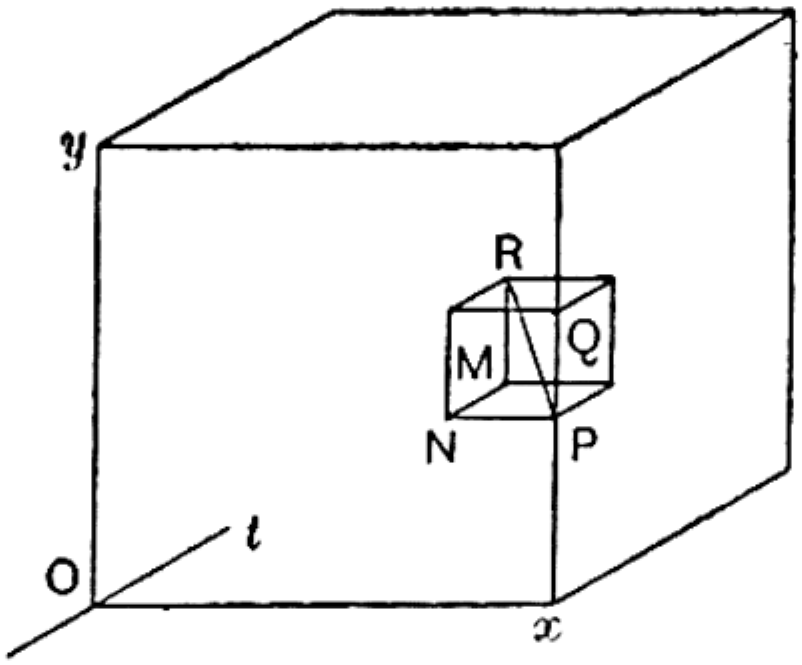

FIG. 76

Thus if we suppose that PO is a vertical pole planted at the earth's equator as shown in Fig. 75, we know that successive rod-lengths marked out along the pole from P radially downwards towards O will indicate successive equal "intervals" and *not* successive equal changes of radius r. The change in *radius* which occurs when one "rod-length" is stepped off will be $dr = \sqrt{\gamma}\,ds$, where ds is unity, because we are speaking of one "rod-length" and the rod *under these conditions* measures units of *interval*. This does not mean that a dweller on earth could not measure changes of radius by means of his rod. From his point of view this would be just what he was doing, for he would see no sense in the γ equation. For himself the earth dweller postulates values for the g's which make his own immediate region free from hummocks; he recognizes nothing of the rod-contraction and clock-slowing which the distant observer sees. In our present discussion we are taking the point of view of the observer looking on from afar.

The equation just given may be written:

$$dr = \left(\sqrt{\frac{r - 2m}{r}}\right) ds$$

and since m is a "constant" the factor multiplying ds is one which becomes smaller as r becomes smaller, which means that each element of interval (ds), namely each "rod-length" means a change in radius which is smaller and smaller as the radius gets smaller. The distant observer expresses this by saying that the rod contracts more and more as the earth's centre is approached.

As r gets so small as to be approaching closely the value "$2m$" the successive marks (rod-lengths) upon the pole PO get closer and closer together until finally they fall each upon the preceding one—the gravitational field having made the rod contract to nothing!

What is meant by the "distance $2m$" from O?

In Einstein's equation "m" is a constant: when it is taken to mean the earth's gravitational mass the equation gives results which agree with experiment.

How can m be expressed in units which are consistent with the other units in the equation?

Experiment led Newton to the law $F = m \times m'/d^2$ expressing the gravitational force between two masses m and m' distant d apart. In this form of the equation the units of mass have been chosen in such a way that the gravitational constant is unity.

Since force is defined as (mass × acceleration) we may write, for the force on mass m' at distance R from the centre of the gravitating mass m,

$$F = m' \times A = m' \times m/R^2,$$

and we see that the acceleration experienced by the mass m' is (m/R^2).

We use the moon as test body; it follows an almost circular orbit of radius $3{\cdot}9 \times 10^5$ kilometres around the earth, and its peripheral velocity in its orbit is about 1 kilometre per second, or in the units of velocity in which light's velocity (3×10^5 km./sec.) is unity, about $1/(3 \times 10^5)$ of a unit.

We know, however, that the acceleration of a body moving with orbital velocity v in a circle of radius R is (v^2/R); hence, for earth $v^2/R = (1/3 \times 10^5)^2/R = A$, and since, from Newton's law of gravitation (with which we know that the moon's motion round the earth is closely consistent) we had $A = m/R^2$, we write, finally,

$$A = m/R^2 = v^2/R = (1/3 \times 10^5)^2/R,$$

and, since $R = 3{\cdot}9 \times 10^5$ therefore $m = 4$ millimetres in kilometre measure.

The steps in the calculation show exactly with what understanding we thus express a Mass in units of Length. We may compare the convenient unit of *distance* used in astronomy, called the *light-year*, or, to take another instance, the expression of atmospheric *pressure* in *millimetres*. In every such case there is an implicit understanding of the conditions.

Applying the result $m = 4$ mm. we have for the earth $(1 - 2m/r) =$ zero when $r = 8$ mm. This means that, if the earth could contract to a radius of 8 mm.—still retaining its present mass—then an earth dweller, seen by a distant observer, should appear to be no more than a bird's-eye view of himself. By climbing up a radial pole he could get further away from earth's centre—and so improve his "figure"—but no efforts to climb *down* the pole would bring him nearer than 8 mm. from earth's centre. The earth would appear to the distant observer as *impenetrable**.

Conservation of mass and of energy

That there must be something fundamental about matter is demonstrated by the facts of the conservation of momentum and of energy. In Newtonian mechanics mass is regarded as the one unchangeable attribute of matter. When a given set of identifiable particles of matter, each possessing some initial velocity relative to a given observer, is left to itself, undisturbed except by mutual collisions of the particles, the velocities of the individual particles will suffer repeated changes, but experiment shows that, with extreme precision, the sum of the products of the masses and their own velocities will remain constant. This experimental fact is remembered by aid of a rule called the "law of the conservation of momentum."

---

* We choose to overlook the difficulty of "observing" within a region of such high refractive index!

In addition it is found that the sum of the products of the masses and the squares of their velocities is constant, and to cover this fact the law of the "conservation of energy" has been invented.

The actual amount of the momentum or of the energy belonging to a given group of particles, differs according to the observer's motion relative to the group, but the conservation of this amount, whatever may be its magnitude for the observer concerned, is true for every observer. The observers must differ, however, in respect only of uniform velocity. The condition that the group of particles shall be "left to itself" means that no forces shall be introduced. If the observer began to change his velocity relative to the system of particles he would regard the particles as suffering acceleration relative to himself, that is to say, as being acted on by forces.

It will now be shown that both these laws are inconsistent —in their Newtonian form— with Relativity.

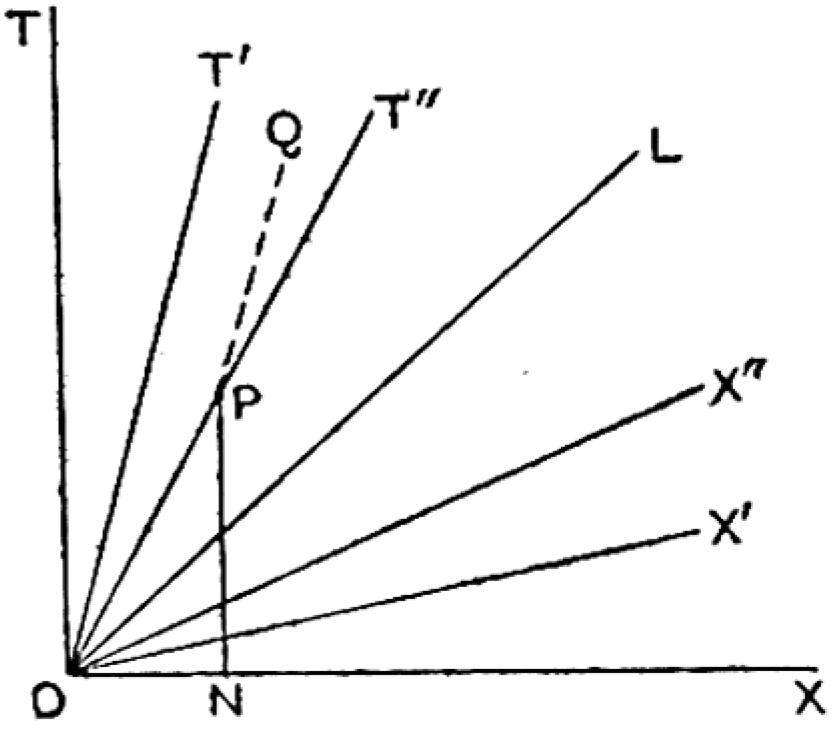

FIG. 77

We consider two particles *A* and *B* of equal Newtonian mass *M*, the one, *A*, at rest relative to *S*, and the other, *B*, moving with half light's velocity in the direction *OX* (Fig. 77). The velocity of *A* is nil and that of *B* is ½, hence

$$MV_A + MV_B = (M/2).$$

Let *A* be at rest at *N* and so have the space-time path *NP*, then *B*, which we assume passes through *O*, must collide with it at *P*. We assume that the two particles adhere on collision, then, since the new velocity of the combined particle, of mass $2M$, must be such that the total momentum remains $(M/2)$, it follows that the new velocity must be $1/4$; that is to say the doublet particle will move along the space-time path PQ parallel to the time axis OT' of an observer S' moving with $\frac{1}{4}$ of light's velocity relative to S. Before the collision S' found A moving with about $\frac{1}{4}$ of light's velocity relative to him and B also with about $\frac{1}{4}$ of light's velocity, the sum of the momenta of A and B being about $(M/2)$. Now, after collision S' finds the combined mass moving parallel to his own time axis OT' and so at rest relative to his frame, the total momentum being thus nil. Clearly then, if the momentum of the group of particles A and B is conserved for S it is not conserved for S'.

Similarly the energy, if conserved for S, was not conserved for S'.

Energy and momentum must receive some new definition if they are to be conserved and, at the same time, are to fit into the scheme of Relativity.

Velocity is a relative quantity: if, therefore, its product with some other quantity is to be absolute this other quantity cannot itself be absolute, or, if the total MV for a system is to be the same for all observers then M must be a relative quantity.

In order to find what modification is required in the definitions of energy and of momentum we must first see clearly why it is that momentum $(M \times V)$ is not conserved for the two particles A and B.

Momentum is defined as $(M \times v)$ for each particle, M being the Newtonian mass and v the velocity x/t, where x is the distance travelled by the particle in time t. It is convenient to write $v = x$, allowing x to stand for the distance travelled in unit time; then the momentum of a particle of mass M is $(M \times x)$.

If two particles A and B move through O in the direction Ox with different velocities, A covering in unit time the distance x_A and B the distance x_B the momentum of the system must be

$$Mx_A + Mx_B = M(x_A + x_B),$$

the space-time paths being OA and OB (Fig. 78). The simplest way of adding x_A to x_B—the points A and B lying both on the mesh-line for unit time—is to draw AD equal and parallel to OB meeting the mesh-line $t = 2$ at D; then $ND = (x_A + x_B)$ and is proportional to the total momentum of A and B.

If, while passing through O the particles had stuck together or had collided with rebound their velocities would have changed and their paths in unit time would have been other than OA and OB, but, for conservation of momentum these paths, when strung together as before, would have reached the same point D. If, for instance, the particles stuck together the doublet would follow the path OE, E being the mid-point of OD and of AB.

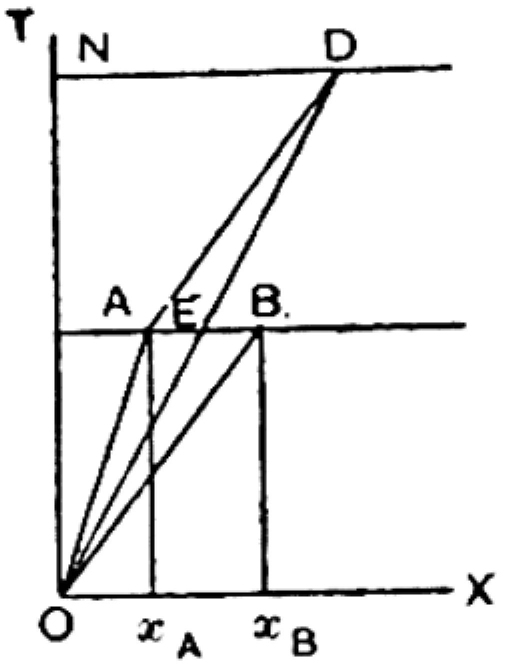

FIG. 78

We see then that, for conservation of momentum the space-time path of the doublet must be along that diagonal of the $OAOB$ parallelogram which bisects the diagonal parallel to OX.

The same condition must hold for any other observer S' having axes $T'OX'$ who is watching the same pair of particles (the world-lines of which, being independent of any observer, must always lie along OA and OB); that is to say, if the momentum is to be conserved also for S' then the path of the doublet must lie along that diagonal of the $OAOB$ parallelogram which bisects the diagonal which is parallel to OX'.

Inspection of Fig. 79, in which OX' is the x axis of S' and OX that of S, shows that the diagonal OE' of an $OAOB$ parallelogram in S' frame cannot lie along the diagonal of the $OAOB$ parallelogram which is in S frame, because while EE' must be parallel to BB' (E and E' being the mid-points of AB and AB') OE is oblique to BB'.

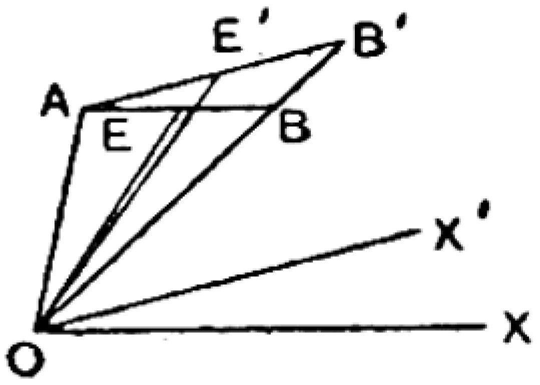

FIG. 79

We conclude, therefore, that if the momentum of A and B

is to be conserved for S (a condition sending the doublet along OE) it cannot be conserved for S' (a condition sending the doublet along OE').

The answer to the question how the definition of momentum may be modified so as to insure conservation for all observers is suggested by consideration of Fig. 79. If all observers drew the same $OAOB$ parallelogram then there could be no ambiguity about the doublet's path; the combined particle $(A + B)$ would assume, in every frame, the velocity required to ensure the conservation of momentum.

The $OAOB$ parallelograms differ because of the differing slopes of the x axes of S' and S, the definition of momentum making it necessary for each observer to put A and B on one of his own time lines (parallel to his own x axis).

Can we define momentum in such a way that AB and AB' coincide?

Referring to Fig. 78 we remember that OA and OB were made to end on the unit time line because this made x_A and x_B proportional to momentum. x_A and x_B had to be the distances covered *in unit time* by A and B.

Unit time means different paper distances to different observers—is there any quantity which means the same, in paper distance, to all observers? "Interval" is such a quantity. Tell S' to mark off unit interval along OA and along OB: the points he marks will agree precisely with those marked off by S or any other observer.

Let us therefore define momentum as the product not of mass into distance covered per unit time (dx/dt) but of mass into distance covered per unit interval (dx/ds); then all observers will draw similar $OAOB$ parallelograms and momentum, thus defined, will be conserved.

Newtonian momentum is defined as $Mv = Mx/t$, the new definition is

$$\begin{aligned}\text{Momentum} &= M'x/s \\ &= M'\,(x/t)\,(t/s) = (M'\,t/s)\,(x/t).\end{aligned}$$

Thus the quantity $(M' t/s)\, v$ is conserved for all observers and we may say that *momentum* is conserved and may still write momentum as $M \times v$ provided the Newtonian definition of mass as an absolute constant be relinquished and M be recognized as a quantity $(M' t/s)$, containing an absolute factor M', but whose value varies.

The new meanings of momentum and of mass

The equation

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2$$

gives $$(ds/dt)^2 = 1 - (dx^2 + dy^2 + dz^2)/dt^2 = 1 - u^2$$

or $$dt/ds = 1/\sqrt{1 - u^2},$$

u being the velocity of a particle in a space-time frame Ox, Oy, Oz, Ot. Hence instead of $(M' t/s)$ we may write

$$M'/\sqrt{1 - u^2} \text{ or } M'\beta,$$

and defining mass M as equal to $M'\beta$ we say that *mass is a quantity which varies with velocity*. With this definition of mass the law of conservation of momentum $(M \times v)$ holds for all observers.

The factor β tells us that the mass of a body varies with its velocity, increasing by the same percentage that the body's length, and therefore also its volume, decrease. Hence the density of a body, defined, as its mass per unit volume is a relative quantity as well as length and mass.

The relation of mass to energy

$\beta = 1/\sqrt{1 - u^2/c^2}$ may be written

$$\{1 - (u/c)^2\}^{-\frac{1}{2}} = 1 + \tfrac{1}{2}\,(u^2/c^2) + \dots$$

hence $$M = M'\beta = M' + \tfrac{1}{2} M' u^2/c^2 + \dots$$

or, putting $$c = 1$$

$$M = M' + \tfrac{1}{2} M' u^2.$$

This means that M (which usually is called the Newtonian mass, although it is not constant) is made up of an absolutely constant part M' and a part consisting of the kinetic energy of the body. Thus a body's mass may be increased by adding energy, and the conclusion seems warranted that energy and mass are two names for one and the same thing.

Relativity accepts the principles of the conservation of Mass, of Energy and of Momentum, provided mass be defined as (Newtonian mass/β).

The variation of mass with velocity

The equation

$$M\,(=\text{Newtonian mass}) = \frac{M'\,(=\text{invariant mass})}{\sqrt{1-u^2}}$$

shows that when $u = 1$, M becomes infinite. This means that if the velocity u of a particle ever should reach 186,000 m.p.s. its mass would become infinite and an infinitely great force would be required to increase the particle's velocity beyond this value; in other words, *the velocity of a massive particle can never exceed light's velocity.*

No velocity can exceed 186,000 miles per second

The statement that no velocity can exceed 186,000 m.p.s. is apt to be misunderstood. Someone asks, "Is it true that the earth rotates relatively to the fixed stars about once in 24 hours, and that therefore—motion being relative—the stars may be regarded, legitimately, as making immense circles around the earth in 24 hours? If so then calculation shows that the peripheral velocity of a star at the comparatively small distance of 10,000 millions of miles is greater than 186,000 miles per second."

This is true enough, and it is not inconsistent with Relativity.

If we measure that star's velocity relative to earth we find it is practically nil—that is why it is called a "fixed" star. "Peripheral velocity" is a convenient term to use in calculations, but a particle may have "peripheral velocity" however great without possessing measurable or real velocity at all. The mind considers a peripheral velocity to be a real velocity—that is to say a velocity which can be measured—because it insists on assigning to the particle a *path fixed in space* relative to which the rotating particle moves—like a cyclist circling a track.

In the statement that no particle can have a velocity greater than c the word velocity means "real" velocity, and a "real" velocity is one which provides the same numerical result of measurement to every conceivable observer. The equations for relative velocity, given on p. 238, shows that no particle's velocity relative to an observer can exceed c.

A favourite instance which seems to refute this is that of a pair of electrons shot off simultaneously in opposite directions from a particle of matter. It is known that matter may emit electrons radially in all directions with velocities as great as 80 % of light's velocity. Thus in Fig. 80 a particle of matter M may send out at the same instant one electron A along MA with velocity $0{\cdot}8\,c$ and another B along MB with equal velocity. *Obviously* then the velocity of B relative to A is ($2 \times 0{\cdot}8\,c =$) $1{\cdot}6\,c$, some 60 % greater than light's velocity.

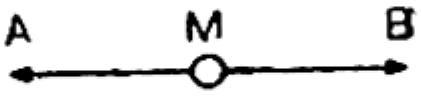

FIG. 80

The *obviousness* is Newtonian and can be admitted only by one who postulates an identifiable rigid background of space in which the motions of A and B are to occur. Strictly the velocity $1{\cdot}6\,c$ is fictitious. That which is real in the velocity relation of A to B must be that which appears the same to every observer; the same, namely, as it appears say to A or to B. The equation for relative velocity shows that to A, as also to B, the relative velocity of the particles A and B is a little less than c and is not $1{\cdot}6\,c$.

It is not inconsistent with experiment to assume tentatively, as Relativity does, that no velocity can exceed 186,000 miles per second.

The motion of Mercury's perihelion

The space-time distortion produced by matter—or which, so far as is known, *is* matter—involves a pure space distortion. Thus if an exactly circular path be described around a planet—a path whose measured distance from the planet's centre is constant—measurement with a rod, or by optical means, will show that the peripheral length of that path is less than 2π times that constant radial distance. Strictly "peripheral length" in such a case is, like "peripheral velocity," fictitious, but in such discussions it is usual to use the Newtonian conception of a space, fixed relatively to the system of "fixed" stars, in which, and relatively to which, the motions of planets occur.

The deviation from π is predicted in Einstein's equation

$$ds^2 = -(1/\gamma)\,dr^2 - r^2\,d\theta^2 + \gamma\,dt^2$$

in which $\gamma = (1 - 2m/r)$.

In measuring the circumference of the path we use polar co-ordinates and consider ds as the hypotenuse of a little triangle whose other sides are $r\,d\theta$ and dr $(= 0)$. The path under consideration may be thought of as a complete ring of wire—a static thing, around which measurement may be made in no time so that $dt = 0$—each little ds triangle lying solely in the plane of r and θ. Then for each ds triangle

$$ds^2 = 0 - r^2\,d\theta^2 + 0,$$

or with the usual change of sign for convenience, $ds = r\,d\theta$.

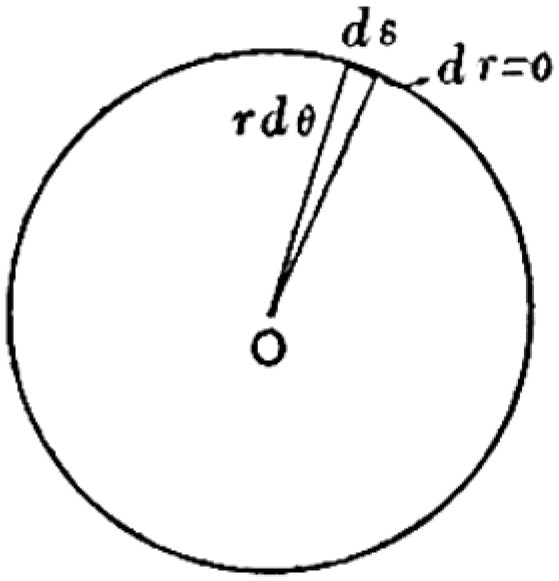

FIG. 81

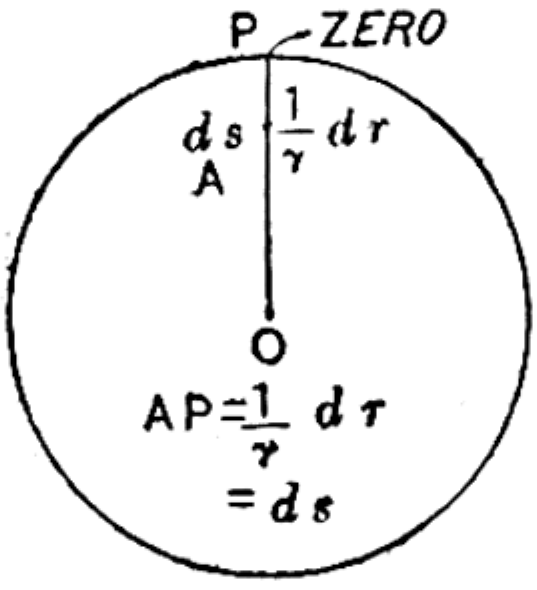

FIG. 82

With a little care we can avoid possible confusion as to what it is—interval or length—which the rod measures in a given case. When an observer sets about measuring the interval between any pair of points P and Q in his TOX diagram he must use a clock for finding dt and a rod for dx; he then writes, $ds^2 = dt^2 - dx^2$ and so finds ds; but in the special case in which P and Q lie both on OT, or any other of the warp lines, he knows that dx is nil and that therefore $ds = dt$ and his clock measures interval. If, on the other hand, both P and Q are on a woof line then the *rod* measures interval.

In Figs. 81 and 82 we are using a system in which only space measurements occur, and a mesh-system whose woof lines are circles concentric with O and warp lines are radial, thus whether measuring radially or circumferentially the rod measures *directly* elements of interval.

Thus the elements of circumference, $rd\theta$ in Fig. 81, are measurable with a rod and the equation just written says that when laid circumferentially the rod measures *interval*, ds standing for the "measured length" of each element of the circumference. The sum of all the elements ds taken around the circumference is the sum of

$$(rd\theta + rd\theta + rd\theta \ldots) = r\,(d\theta + d\theta + d\theta \ldots) = r \times 2\pi$$

since, in going right around the circle the sum of the little angles turned through by the rod as it passes from one ds triangle to the next is the total angle 2π radians. Thus the measured length of the circular path is $2\pi r$.

In measuring the radius of the path, the ds triangles have hypotenuse ds and sides $(1/\gamma)\,dr$ and zero, and the equation is $ds = (1/\sqrt{\gamma})\,dr$ and as $\gamma\ (= 1 - 2m/r)$ is always less than unity it follows that ds (the quantity the rod measures) is always greater than dr (that is to say, the rod contracts). Hence the sum of all the measured elements ds stepped off radially by the rod, is greater than r.

We conclude that the geometry of the circular path is such that the measured circumference is $2\pi r$ and the measured radius *greater* than r, that is to say the path is a circular one whose measured length is less than 2π times the measured radius.

We apply this fact to the approximately circular path of Mercury or any other planet round the sun. The radius of the path has long been known by optical measurement: let it be called the measured radius R. According to Newtonian principles when Mercury goes once around the sun it traverses an orbit equal to $2\pi R$. If, as Einstein's theory demands, the space around the sun is non-Euclidean and Mercury's measured path is less than $2\pi R$ for one complete angular revolution, it follows that in travelling an orbital distance $2\pi R$ Mercury must make more than one complete angular tour around the sun: the beginning and the end of the orbit, of length $2\pi R$, will overlap and Mercury will get gradually ahead of the position which, on Newtonian principles, it ought to occupy.

Actually Mercury's orbit is elliptical and the gradual advance just spoken of shows itself in a gradual rotation of the elliptical figure of the orbit.

Outline of a method of deduction of Einstein's law of gravitation from the Lorentz Transformation

As the field due to a spherical mass is symmetrical, it will suffice, in deducing the law of gravitation, to consider the ds^2 equation for one plane containing the centre of the mass.

For convenience polar co-ordinates (slightly distorted by the gravitational field) will be used.

Experience says that space around the sun is very nearly Euclidean, hence the g's of the ds^2 equation for the region of the sun must be such, very nearly, as to make the equation read

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2,$$

or, as we shall use r for x and θ for y

$$ds^2 = dt^2 - dr^2 - d\theta^2,$$

the dz falling away because we are considering only a 2-dimensional plane.

Actually the equation will be

$$ds^2 = -g_{11}dr^2 - g_{22}d\theta^2 + g_{44}dt^2$$

in which, if we make $g_{22} = r^2$ (see Chap. IX) then g_{11} and g_{44} must be nearly unity: it is the values of these that we have to find.

As a preliminary we consider again the results deduced on pp. 244 *et seq.* from the Lorentz Transformation, in regard to the effect of gravitational fields upon rods and clocks.

We found that in a field of potential ϕ, a unit rod became $(1 - \phi)$ units long and a seconds tick took $(1/1 - \phi)$ second, where ϕ the gravitational potential $= km/r$ (see p. 247). We employ units in which k is unity, taking m in gravitational units of mass, and we have

$$(1 - \phi) = (1 - m/r),$$

m is here the invariant mass, not the Newtonian mass, although so small is the difference, that we define m/r as the gravitational potential in the Newtonian sense.

First we note that, as m/r is usually very small compared with unity, its square may be neglected in comparison with unity, and we have

$$(1 - m/r)^2 = (1 - 2m/r + m^2/r^2) = (1 - 2m/r) = \gamma$$

and $$\sqrt{1 - 2m/r} = (1 - m/r) = \sqrt{\gamma}.$$

A distant observer watching a sun man measuring light's radial velocity with his shortened rod and slow-going clock, would expect the sun man to find it *greater than unity*. The expected value would be, on account of the short rod, $\{1/(1 - \phi)\}$ instead of unity, and this, on account of the long second should be raised to

$$\{1/(1 - \phi)\}\{1/(1 - \phi)\} = 1/(1 - \phi)^2 = \{1/(1 - 2m/r)\} = 1/\gamma.$$

For transverse rays the onlooker would expect the sun man to find the value $1/\sqrt{\gamma}$ instead of unity, for only the clock is then affected (a rod held parallel to the sun's surface suffering no contraction).

We know, however, that the sun man *must* find the value of light's velocity to be unity in each case, hence, we, the onlookers, have to postulate that light travelling radially suffers a reduction of velocity in the sun's field, namely it has

$$\text{Radial velocity} = (dr/dt) = \gamma,$$

$$\text{Transverse velocity} = (rd\theta/dt) = \sqrt{\gamma}.$$

It was shown on p. 251 that light travels along space-time paths of zero interval length, hence, for light, $ds = 0$ and the ds^2 equation is

$$0 = -g_{11}dr^2 - r^2d\theta^2 + g_{44}dt^2.$$

For a ray travelling radially ($d\theta = 0$) we have, then

$$0 = -g_{11}dr^2 + g_{44}dt^2.$$

$$\therefore\ (dr^2/dt^2) = g_{44}/g_{11} = \gamma^2 \text{ (as above)},$$

and for a transverse ray ($dr = 0$)

$$0 = -r^2 d\theta^2 + g_{44} dt^2.$$

$$\therefore\ (r^2 d\theta^2 / dt^2) = g_{44} = \gamma \text{ (as above).}$$

Finally, therefore, since $g_{44} = g_{11}\gamma^2 = \gamma$ we have $g_{11} = 1/\gamma$ and the law of gravitation reads

$$ds^2 = -(1/\gamma)\, dr^2 - r^2 d\theta^2 + \gamma dt^2,$$

where $\gamma = (1 - 2m/r)$.

Is space infinite?

According to common-sense as well as to classical mechanics space is infinite, but the suggestion of Einstein and other Relativists is that space may be finite though unbounded.

If finite, space must be of conceivably measurable volume and the mind asks what can there be, except more space, beyond the limits of that volume?

The difficulty arises from the essentially different points of view taken by a man as a "human individual" and as a "scientific observer." The former's world is made up of all things he can conceive, the latter's is confined to those which he can measure. The former conceives that if he had the necessary vehicle he might go out in any chosen direction through space and might keep on travelling for ever, avoiding, by proper choice of direction, any return to regions already traversed. The scientific observer is concerned not with what conceivably might be but with what actually is. He watches the journey through space of the most space-searching thing he knows, namely, a ray of light, and finding its behaviour consistent with the assumption that space is finite, while no known phenomena demand that it shall be infinite, he adopts the former assumption.

There may exist space beyond the regions which light is able to reach but there is no evidence of its existence.

Matter attracts light, and though it may allow a ray to encircle the material universe it does not allow it to escape.

Alternative suggestions in regard to the finite but unbounded nature of space, or of space-time, are expressed mathematically, but cannot be explained in terms of ordinary conceptions.

Some light may perhaps be thrown upon one of these suggestions by referring first to another of the "absurdities" maintained by Relativity, namely, that matter, of sufficient density is impenetrable.

We suppose that a man with a yard rod is at the top of a pole planted radially in a dense sphere of matter, and begins to measure the pole from its top downwards. The increasing gravitational field causes the rod to contract more and more as the man climbs down the pole, so that the progress made towards the mass's centre for each rod length stepped off becomes ever less. So dense is the matter that just as the man reaches its surface the rod's length becomes zero; he therefore goes on stepping off rod lengths for ever but gets no nearer to the centre of the mass. Science concludes that the mass is impenetrable, but common-sense maintains that the idea of impenetrability is absurd.

Relativity has shown that although the problem of making an infinite space compatible with known phenomena appears impossible, one of the kinds of space which are compatible is a space which would exhibit the following peculiar measure relations. It is known as "spherical" space but must not be thought of as sphere-shaped.

Concentric with any chosen point in space, say, for instance, earth's centre, we imagine an unlimited number of shells, like the skins of an onion. By means of his standard rod an observer sets about measuring the area of each successive shell. He finds at first that successive shells have successively greater areas, but at length he reaches a shell of maximum measured area, after which all succeeding shells prove to have diminishing areas, until ultimately a shell is reached having an immeasurably small area. The observer has reached the "antipodes" of space.

Such space is known as "spherical" space by analogy with the 2-dimensional space forming the surface of a sphere. If, on such a surface, I set out to measure, from a given centre or pole, the periphery of successive circles I come at length to the equator—the maximum circle—after which succeeding circles have ever diminishing peripheries until, finally, at the antipodes of my pole, I encounter a circle whose area is nil.

Having reached the antipodes of "spherical" space there simply is no more space to measure. Common-sense says, measurable or not I *know* there must be unlimited space in which your weird universe of finite space floats. To this dogmatic position there is no answer except to say that to science, things which cannot be measured have no "real" existence.

Common-sense may say, if it likes, that the "reason" the areas of successive shells prove to have "wrong" dimensions is that, in such space (for some reason not expressed in the Lorentz equations), the observer's rod expands as the observer reaches more distant shells, so that the areas "appear" smaller than "really" they must be, until finally, the observer's rod having become very long, all succeeding shells appear to be infinitely small.

It is scientifically meaningless to postulate existence for a thing which is not either observed or at least observable, and to science the question "how much space exists?" means "how much space is explorable?" Radiation is the most thorough instrument of exploration known to science, hence if the extent of space explorable by radiation is finite the conclusion is warranted (tentatively) that space itself is finite. Conversely, to say that space is infinite is scientifically meaningless unless we have shown space to be explorable without limit. If, as experiment indicates, radiation does not travel out to infinity, there is nothing within our knowledge that can. If radiation did indeed go out to infinity our material universe would be gradually dwindling, for radiation is energy and energy probably is equivalent to mass. Such diminution would be annihilation and would imply that the material universe contains nothing that is permanent.

The future development of Relativity

Though embracing gravitation the General Theory of Relativity has not yet succeeded in embracing electro-magnetic fields of force; this makes the theory in its present form incomplete though not inaccurate. The work now being done by Einstein and Weyl promises to bring all phenomena into the theory. Just as the Theory of Relativity in its present form has resulted simply from the removal of unnecessary and un warranted hypotheses from classical science—namely the hypotheses that lengths and times were independent and unambiguously measurable quantities—so the further development of Relativity is being based upon further removal of unwarranted hypotheses. The theory in its present incomplete form assumes that there exist unambiguously measurable elements called space-time intervals, but this assumption is based upon an earlier one now seen to be unwarranted, namely, that having equipped all observers with rods and clocks which, at the outset—close together and relatively at rest—are in agreement, these standards will remain (subject, of course, to the FitzGerald and gravitational contractions) correct. There is no strict warrant for this assumption—that only its present state of motion affects a rod or a clock—the space-time *career* which a standard has followed may possibly affect it as well.

The FitzGerald and gravitational "effects" upon standards may be spoken of as the "causes" (to use a word which, though convenient, has no place in the universal observer's vocabulary) of Gravitation and of all the phenomena of mechanics, but they do not account for electro-magnetic fields of force. If the standards undergo changes depending on their past careers, then it is to be expected that these changes may prove to be the "causes" of the phenomena of electro-magnetism. In order to allow for such possible changes it has been found necessary to introduce four quantities additional to the ten g's in the general ds^2 equation. These four additional quantities are found compatible with the four fundamental equations of electro-magnetism (see Eddington, *Space, Time and Gravitation*, p. 172. Camb. Univ. Press, 1920).

CAMBRIDGE: PRINTED BY
W. LEWIS
AT THE UNIVERSITY PRESS